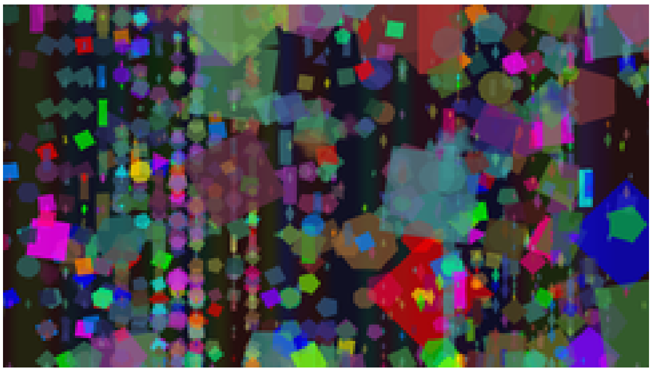
TÉCNICO LISBOA



From Music to Image
a Computational Creativity Approach

Luís Ricardo dos Santos Aleixo

Thesis to obtain the Master of Science Degree in

Information Systems and Computer Engineering

Supervisors: Prof. Helena Sofia Andrade Nunes Pereira Pinto
Prof. Nuno Manuel Robalo Correia

Examination Committee

Chairperson: Prof. João António Madeiras Pereira
Supervisor: Prof. Helena Sofia Andrade Nunes Pereira Pinto
Member of the Committee: Prof. Carlos António Roque Martinho

November 2020

Acknowledgments

In the first place I would like to thank my mother Ana, my father Luís, and my sister Catarina, for being there every time I needed, every time I thought I could not do it, every time I wanted to give up. They always gave me words of encouragement, caring, and belief in the right moments, providing all the extra help I needed at home during all these years. I also would like to thank my grandparents for the support they gave me no matter what. A mention for the rest of my family to their help since the very beginning of my journey at Instituto Superior Técnico.

I also would like to thank all my friends, both from IST and Évora, my hometown, who were always ready to help whenever needed. A special mention for two friends, Francisco Centeno and Edgar Meireles. They were always there for me, during the good and bad times, supporting and encouraging me every day since the very beginning. Their daily strength, companionship, ideas and beliefs on my work were essential to reach this point.

Last but not least, a special word for my advisors, H. Sofia Pinto and Nuno Correia, who were always there, supporting my thoughts and ideas that has made this Thesis possible. I also would like to thank professor Rui Jorge for the extra help he provided regarding the musical domain. Without their insight, support and sharing of knowledge, I would have never reached the end.

To all, and each one of you who believed in me - Thank you.

Abstract

Creativity is an essential factor when creating something new or solving problems. It is directly linked with inspiration, that may come from within us, or from the surrounding environment. Computational creativity is a research field that focuses on the replication, simulation and modeling of creativity using machines. Some attempts have been made to model and implement inspirational systems that rely on cross-domain associations between domains.

In this Thesis, we propose a possible approach for a cross-domain association between the musical and visual domains. We present a system that generates abstract images having as inspiration music files as the basis for the creative process. The system extracts available features from a MIDI music file given as input, associating them to visual characteristics, thus generating three different outputs. First, the Random and Associated Images - that result from the application of our approach considering different shape's distribution - and second, the Genetic Image, that is the result of the application of one Genetic Algorithm that considers music and color theory while searching for better results. The results of our evaluation conducted through online surveys demonstrate that our system is capable of generating abstract images from music, since a majority of users consider the images to be abstract, and that they have a relation with the music that served as the basis for the association process.

We believe that the proposed system provides an interesting solution for possible associations between these two domains. However, work on the topic can still be done to improve the quality of generated images, as well as the assumptions made over the development.

Keywords

Computational Creativity; Music Analysis; Image Generation; Cross-Domain Associations; Genetic Algorithm

Resumo

A Criatividade é um factor essencial na criação de algo novo, ou na resolução de problemas. Está directamente ligada à inspiração, que pode surgir de dentro de nós, ou do ambiente envolvente. A Criatividade Computacional é um campo de investigação que se concentra na replicação, simulação e modelação da criatividade utilizando máquinas. Algumas abordagens têm sido desenvolvidas para modelar e implementar sistemas inspiracionais que se baseiam em associações inter-domínio.

Propomos uma possível abordagem para uma associação entre os domínios musical e visual. Apresentamos um sistema inspiracional que gera imagens abstractas tendo a inspiração em ficheiros de música como base para o processo criativo. Desta forma, o sistema extrai características de um ficheiro MIDI dado como input, associando-as a características visuais, gerando três tipos de outputs. Primeiro, a Imagem Aleatória e a Imagem Associada - que são um resultado directo da aplicação da nossa abordagem considerando diferentes distribuições de formas - e segundo, a Imagem Genética, que é o resultado da execução de um Algoritmo Genético que tem em consideração teoria musical e harmonia de cores na procura de melhores resultados. No final, os resultados de uma avaliação feita através de inquéritos online demonstram que o sistema é capaz de gerar imagens abstractas que têm uma relação com a música que serviu de base para o processo de criação.

Acreditamos que o sistema proposto oferece uma solução interessante para possíveis associações entre estes dois domínios. Contudo, existe ainda trabalho que pode ser desenvolvido para melhorar a qualidade das imagens geradas, bem como as associações estabelecidas no desenvolvimento.

Palavras Chave

Criatividade Computacional; Análise Musical; Geração de Imagem; Associações Inter-Domínio; Algoritmo Genético

Contents

List of Figures

List of Tables

Acronyms

AI	Artificial Intelligence
CC	Computational Creativity
GA	Genetic Algorithm
IGP	Interactive Genetic Programming
CGP	Cartesian Genetic Programming
ANN	Artificial Neural Networks
GAN	Generative Adversarial Networks
CAN	Creative Adversarial Networks
EMI	Experiments in Musical Intelligence
VIV	Visual Information Vases
BPM	Beats Per Minute
MIDI	Musical Instrument Digital Interface
IQA	Image Quality Assessment

1 Introduction

Contents

Humans have been using tools to improve or extend their capabilities. The discovery of the wheel that dates back to 3500 b.c., the use of artifacts built to fight against nature and survival, or the paintings in caves' walls during the stone age, are some examples of this capability. This may prove that humans have been capable of solving problems with creative thinking since immemorial ages, even though the concept was introduced much later in the time.

In most Ancient cultures, the creativity concept did not exist. In Ancient Greece, for example, Art was seen as a form of discovery and imitation, not creation. Its first notion arrived in Western cultures through Christianity, as a matter of divine inspiration - the belief that some entity above helps on the creative moments. However, the modern sense of creativity was first seen during the Renaissance, where the term was first applied to humans [24]. Since that time, many theories have emerged about its definition, changing over the history, as has the term itself.

Rhodes, M. [40] in his attempt to define creativity introduced the four P's of creativity:

- **Person** - refers to those who create.
- **Process** - refers to the chain of actions that take place in creative work.
- **Press** - refers to external environment pressure that impacts creative process.
- **Product** - refers to the result of the creative work.

These concepts can be combined together to loosely build a definition of human creativity: a **process** in which a **person** generates a **product**, having in consideration the **press** from the external environment. However, since we are considering the creative process, we are particularly interested in the press notion. Press is everything that influences humans while having creative moments, closely related to the inspiration concept, defined[1] as *"the process of being mentally stimulated to do or feel something, especially to create something creative"*. It is something that arises and motivates us to do all kinds of tasks. By having inspiration, which can come from inside or from the world that surrounds us, we can generate something new or solve a problem. It may depends on the culture we belong to, the time of the history, or if we are doing it individually or in a group.

Inspiration, as well as creativity, is hard to define in objective terms, since there is no established theory that explains it in objective terms. An essential part of the creative process is to be conscious of the prior work, and to internalize the symbols and conventions of the domain in which it occurs: creativity results when the creator combines previous existing elements and generates somehow new combinations. In the end, the results of a creative process may be tangible or intangible items, such as paintings, ideas, jokes, or scientific discoveries.

The research on Artificial Intelligence (AI) has been transforming the humans' ability to create new ideas, opening new possibilities never taken into account before. Regarding arts and science, researchers have been searching how to build creative software, either as a tool used to improve or

[1] https://www.lexico.com/definition/inspiration

enhance human creativity, or as an autonomous artist, writer, designer, engineer or scientist, following and taking part in the creative process responsibilities. This way, Computational Creativity (CC) emerged as a sub-field of AI, exploring the machine's ability to generate human-level creative artifacts, hence it focuses on the development of software that can be considered creative by humans. In other words, this research area has the goal of replicate, simulate, or model creativity using a computer to achieve one of considerable goals [9]:

- The development of systems capable of human-level creativity.
- The research on algorithmic perspectives on creative behaviors in humans.
- The construction of programs that enhance human creativity, without necessarily being creative themselves.

This field of study concerns itself in the parallel work on theoretic and practical issues in the study of creativity, with one strand of work informing the other - it studies the nature and proper definition and creativity, and implements the systems that exhibit it [9].

Some systems in the area of CC rely on the press concept to translate the creativity and essence of one artifact into a new one. For example, Teixeira, J. and Pinto, H. [46] in 2017, described an inspirational system that establishes a possible analogy between the visual and musical domains - Section 2.3.7. A similar approach was done by Santos et al. [44] in 2020. Braga et al. [5] in turn, proposed a system in 2019 that composes music inspired in sculptures, building a non-strict association between two different domains. In these inspirational systems, the goal is to translate the essence of one artifact into a new one, even if different domains are considered, thus creating Cross-Domain Associations.

A possible approach to this association process is to take features from one artifact to create a new one depending on the creator's objectives or personal choices. The results may be novel and valuable [4], and may be considered creative. However, there might exist unexpected issues, such as essential information from one domain that can be lost, or bias from the person who is creating the associations.

1.1 Motivation and Objectives

Throughout time, much criticism has emerged in the field of computational creativity, such as that artificial creators cannot model emotions, motivations, or selection decisions, or that creativity cannot be algorithmic [34]. However, in a very concrete manner, when humans listen to music, they are processing sound waves, while when watching images, they are processing matrices of pixels.

In the creative process, humans search for inspiration. It may come as a result of inner processes or from the surrounding environment. Depending on the problem, inspiration may come from a different domain. As in most art productions, inspiration plays a significant role - in our work, we use it to associate features from different domains. There are systems that autonomously generate images based on previously trained models, such as the *ETC* [18] or *The Painting Fool* [6]. There are also systems that compose music inspired by artifacts from other domains, such as paintings [46], or sculptures [5]. These systems served as a motivation to ours, since they explore topics related to cross-domain associations.

In this Thesis, we wanted to develop a system based on the inspiration concept. Having as motivation processing music and image algorithmically, our main contribution is the development of a system that exhibits creative behaviour through inter-domain associations, by generating abstract images inspired in musical artifacts. More specifically, we are not only concerned about possible associations between different domains, but also concerned about the usage of mathematics to analyze music and generate images. We consider abstract images as visual artifacts that do not to reflect or convey anything ”concrete” or ”real”. There are different approaches we can consider - from visual representations of musical scores to generation of abstract images through associations one-to-many and many-to-one. However, considering that we are not interested in visual representations of music, as in the case of a sheet music, through the process of association between different domains, we tried the most to avoid associations one-to-one, mapping in various ways as many musical and visual features as possible.

From an input file containing information about the music, the system can consider most of its features - from melody, harmony and rhythm - as inspiration to generate abstract images, emphasising lines, shapes, colors, textures, among others. As well as with humans in their creative process to generate art, it is interesting to see how musical characteristics can be mapped to create abstract images, having the inspiration concept as a basis of the creative process.

At the beginning of this project, we started by considering the following:

- The source and type of music files that would be used.
- What are the essential elements of an image that would be considered.
- Which musical features we would map to the visual domain.
- Which algorithms would be used to provide a way to search for better results.

By answering the previous issues, we had enough information to analyze both domains, and extract all the needed information to make the possible associations. Over the process, features from both domains were chosen according to **logic**, **theories**, and our **aesthetics**. Since the creation of abstract images does not require any specific rules, we do not have any constraints related to the image generation. However, regarding the musical domain, since we want the generated images to fit on a screen, and image size depends on the length of the input file, we limited our input files to music that lasts between three and five minutes. We decided to use Musical Instrument Digital Interface (MIDI) files that already have all the music computationally processed in such a way that it is easier to retrieve and manipulate most of its features. This way, every music that has a correspondent MIDI file can be used as an input for our system. Our dataset is composed of diverse music from different periods and composers that belong to the Classical style due to its diversity and composition of sophisticated forms of instrumental music. However, our approach is independent of the music style, thus any could be considered. Finally, even though we defined a default value for the size of the generated images - 1600$\times$2700 pixels - these dimensions are not strict, since we allow users to define all the desirable dimensions at the beginning of the run of the system. Besides, there is no limitation regarding the screens where the images are displayed.

Our main contribution is one possible association between two different domains, computationally generating abstract images that are inspired from music. Throughout the development of our system, we went through several phases, from research on music theory to the study of color harmonies, shape assembling, and image generation techniques. In the end, to improve the aesthetic value of the generated images, as well as to search for better results, we went through research on Genetic Algorithms, implementing one from scratch. It should be noted that we label our results as "abstract" images, however opinions may vary from person to person. As such, throughout this Thesis, we will call our results as images, or visual artifacts.

Regarding the final results, the system delivers three different images for each music. The first, called the **Random** Image, was generated assigning a random shape and texture to each instrument found on the music file. The second, called the **Associated** Image, was generated assigning a predefined association between musical instruments and respective shapes. It is important to note that several different associations were tried throughout the development of the system, and what we have done is merely one possible approach. To help in the process, we developed one online platform to study and customize all the possible associations between artifacts, available on `https://creativity-2020.nw.r.appspot.com`. Regarding the third version, called the **Genetic** Image, it results from the execution of the Genetic Algorithm that receives the two previous versions to generate the initial population. Its evolution mostly considers color and music harmony between the elements in the canvas, their locations, and shape's distributions.

Currently, we have evaluated four music files, from different periods of time and composers, and the respective translated versions:

1. Georg Handel - *Concerto No. 1 - 2. Largo*
2. Georg Handel - *The Royal Fireworks Suite - 3. La Paix*
3. Wolfgang Mozart - *Symphony No. 40 - 3. Menuetto and Trio; Allegretto*
4. Ígor Stravinsky - *The Firebird - 1. Introduction*

We decided to assess these pairs of music-images through online surveys. In total, 93 people were questioned. The evaluation had good results since the majority of the participants consider the images to be abstract. Besides, however depending on the generated images, the participants believe that there is a relationship between the images and the music. Finally, even though it was not our main objective, it is also worth mentioning that all generated images were generally liked. Having these interesting results, we believe that our system achieved its goals.

Regarding creativity, considering that human opinion has to do with temporal, cultural, motivational and emotional factors (considerations that machines have almost nothing to say), we never know until which point one artifact produced can be considered creative. Furthermore, since it is the programmer who maps the direction between domains, one artifact may easily be considered one avatar of its creator. Nonetheless, we want to contribute to the Computational Creativity field by developing, through cross-domain associations, a system that generates abstract images having as inspiration music files.

1.2 Document Structure

In Chapter 2, we provide background knowledge related to the state of the art for our field. In Section 2.1, we discuss some basic concepts related to Computational Creativity, such as inspiration, creativity, and its types. In Section 2.2, we explain some techniques, presenting their advantages and disadvantages. Furthermore, in Section 2.3, we present some inspirational systems, describing their main contributions to our work. Finally, at the end of this Chapter, we present one analysis that summarises all the studied systems.

In Chapter 3, we present basic background on the musical domain, describing definitions related to music theory in Section 3.1. Explanations on each of the considered features, crucial to the understanding of our work, are presented in the following subsections.

Chapter 4 explores the visual domain by presenting some definitions in Section 4.1. Through this Chapter, we also present detailed research regarding different characteristics of drawings, color schemes, and shape assembling.

Then, in Chapter 5, we describe how we decided all the associations between both domains. In Section 5.1, we explain in detail the whole rationale behind our approach, describing each step we took to reach the final result. In Section 5.2, we describe how we implemented the system, explaining all the modules in the architecture.

In Chapter 6, we present some case studies: in Section 6.1 we detail our dataset, in Section 6.2 we explain the system's process and output, while in Section 6.3 we describe the taken approach to evaluate our system. Then, in Section 6.4, we present and explain the results.

Finally, in Chapter 7, we present the conclusions of our work, as well as possible future work.

2

Related Work

Contents

Diverse research work has been pursued in the Computational Creativity field. In this Chapter, we present some definitions, techniques, and developed systems to provide a background in how humans' creative capabilities are emulated in computational systems to produce novel results.

2.1 Definitions

To define creativity is not an easy task. According to Boden, M. [3], creativity is not a special "faculty", nor a psychological property confined to a tiny elite, but rather a feature of human intelligence, generally linked in everyday capacities, such as the association of ideas. In 1998 [3], Boden considered that a creative idea is one that is not only novel but also surprising and valuable in terms of interest, usability and beauty. Later on, the author redefined her definition of creativity, writing that it can be defined as the ability to generate **novel** and **valuable** ideas [4]. The novelty concept has to do with:

- The mind of the creator only: P-Creativity (P for psychological).
- The mind of the creator, and the whole of the previous history: H-Creativity (H for historical).

Artificial Intelligence should concentrate on the former type [3], building P-Creative artifacts new to the person or system that is generating it, no matter if others have ever had that idea before. Boden [3] further explains this definition, arguing that true creativity results from *transformation of conceptual spaces*, thus identifying three types of creative processes:

- **Combinational Creativity** that involves novel combinations of simple ideas in a well know conceptual space, generating poetry, imagery, jokes, or analogies between different domains, wherein the new ideas share an inherent common structure.
- **Exploratory Creativity** that further explores the potential of structured conceptual spaces, never leaving its bounds. This model is the most successful of the three creative processes, not because it is easy to reproduce, but because it requires considerable experiences in the domain, and analytic power to define the conceptual space and specify the procedures that allow its potential to be explored [3].
- **Transformational Creativity** assumes that instead of merely exploring the conceptual space, the individual does not follow the rules and explores beyond its boundaries. The discoveries transform some dimension of the space *"so that new structures can be generated which could not have arisen before"* [3].

Even though it hardly helps on the discussion in the current format, the following examples of each model of creativity are used by Boden. The *Copycat* system [21], for instance, was introduced as an example of Combinational Creativity because it looks for analogies between alphabetic letter-strings.

BACON, a system that is capable of significant scientific discoveries [27], is given as an example of Exploratory Creativity since it models scientific discoveries using pre-programmed heuristics through the exploration of conceptual spaces. Lastly, the *Automated Mathematician* [28], is an example of Transformational Creativity because it uses algorithms that generate and modify Lisp programs, which were then interpreted as defining various mathematical concepts.

Boden differentiates Combinational from Exploratory Creativity giving as an example the *Jape* system [2] that, given a set of rules, creates jokes by exploring a space of possibilities. Although it seems to be a clear example of Exploratory Creativity, this program is used as an example of Combinational Creativity because the joke-type is made of different combinations of rules and words used in the same conceptual space.

Each of the previous types of creative models can be modeled using AI. However, most of the computer programs use Exploratory Creativity, since it only requires the specification of the rules of the relevant thinking style, providing it to a computer program [4].

In turn, Sawyer, R. in 2011 [45] explains that most of the theories about how the creative process works agree on four basic stages:

- **Preparation** - collecting data, searching for related ideas and listening to suggestions.
- **Incubation** - the previous material is internally elaborated and organized.
- **Insight** - the subjective experience of the *"Eureka!"* moment.
- **Verification** - the evaluation of the insight and elaboration into its complete and formal form.

Although the study of these stages helps us understand the creative process, some are still not well understood on how it occurs. The **incubation** stage, for instance, is often below the surface of consciousness, and it is the least understood stage in the process - previous ideas combine, and the insight stage occurs when certain combinations merge into consciousness [45].

Colton, S. [7] considers that to be considered creative, one software must exhibit behavior that can genuinely be called *skillful*, *appreciative* and *imaginative* - if the software has no skill, it will not be able to create anything valuable; if it has no appreciation of what it is doing, it will not ever understand the value of its work; and if it has no imagination, at best they would only produce pastiches of other people's work.

Koestler, A [26] explains a general theory of human creativity, presenting both the similarities and differences of creativity in three different domains that have no boundaries between them: **Humor**, **Science**, and **Art**. To explore it, the author also introduces two terms to characterize the creative process: *matrices of thought* and *bisociation*. The former represents an ability, habit, or skill that describes the code of "any pattern of activity" governed by fixed rules, somehow flexible. The latter makes a distinction between the routine skills on a single matrix and the creative act - that operates in more than one matrix [26].

This way, the bisociation concept stands for the act of simultaneously using two or more matrices of thought. Thus, the incubation period in the four different stages of the creative process involves the use of different matrices of thought in a *"flexible manner on many different levels of consciousness"* [38].

The author explains in detail each of the three previous domains by making use of the terms described before: *bisociation* and *matrices of thought*:

- **Humor** - Koestler considers that the reaction to something that is deemed to be humor, creating the comic effect, is the consequence of the collision of two matrices of thought. To explain the bisociation act in this domain, the author wrote the following joke:

  "Chamfort tells a story of a Marquis at the court of Louis XIV who, on entering his wife's boudoir and finding her in the arms of a Bishop, walked calmly to the window and went through the motions of blessing the people in the street. 'What are you doing?' cried the anguished wife. 'Monseigneur is performing my functions,' replied the Marquis, 'so I am performing his. '" [26] (pp. 96).

  In this joke, the adultery's context is suddenly bisociated with that of *"the division of labour..."*, breaking the line before the climax. This occurs by replacing the current matrix with another, which seems to be incompatible and unexpected according to what outcome people were expecting, creating the comic effect.

- **Science** - According to Koestler, Discovery, or Science, results from the *integration* of two matrices that were unrelated before. To explain his theory, Koestler cited Archimedes' theory for discovering the volume of non-regular solids using water:

  "No doubt he had observed many times that the level of the [bath] water rose whenever he got into it; but this fact, and the distance between the two levels, was irrelevant to him – until it suddenly became bisociated with his problem. At that instant, he realized that the amount of rising of the water-level was a simple measure of the volume of his own complicated body." [26] (pp. 105).

  When Archimedes stopped all conscious efforts at solving the problem at hand, his subconscious got into a common-sense approach to conclude that the level of water increases when something is submersed on it, enabling the discovery of the measurement of the volume of non-regular solids.

  Koestler's discussion on scientific discoveries raised some interesting concepts:

  - **Ripeness** - when discoveries show some relations to previous knowledge.
  - **Unconscious** - the *"Eureka!"* point is reached when people are not thinking on the scientific matrix, as happened with Archimedes in his discovery.
  - **The Evolution of Ideas** - Koestler explains that the historical progress of science results of different scales of individual discovery, from the periods of preparation and interpretation of the collected data, to the synthesis of a new collective orthodoxy [38].

- **Art** - According to Koestler's view, the aesthetic experience is due to the "*juxtaposition* of two matrices", by bisociation. As an example, Koestler argued that rhythm, rhyme and assonance allow one poet to have creative thoughts by "bisociating sound and sense, metre and meaning". This example shows that there is no unexpected jump, since we are just navigating through the two matrices at the same time - such as in humor.

These three domains - Humor, Science and Art - can be represented in an ordered spectrum without defined boundaries among them. However, the emotional climate may be different in each panel: while humor can stand for aggressiveness (in a comical sense), scientific discoveries can be emotionally neutral, and art artifacts sympathetic and admiring, inspired by a positive kind of emotion. The strategy of the creative process is the same in these three cases: "discovery of hidden similarities" [26].

In the end, the author describes the importance of creative thought in humans writing:

"The creative act, by connected previously unrelated dimensions of experience, enables him to attain a higher level of mental evolution. It is an act of liberation – the defeat of habit by originality." [26] (pp. 96)

Regarding the evaluation on the creative capabilities on computational systems, Ritchie, G. [42] explains that to make judgments about *how creative a computer program is* requires one formal and explicit criteria to assess its results, being guided on how the term is used when applied to human creative behavior. This way, a methodological work was presented with formal characteristics of programs' construction and output, important to evaluate if they behave creatively:

- **Novelty** - *To what extent is the produced item dissimilar to existing examples of that genre?*
- **Quality** - *To what extent is the produced item a high-quality example of that genre?*
- **Typicality**[1] - *To what extent is the produced item an example of the artefact class in question?*

Considering P-creativity items, these three measures may define if one is deemed to be creative [42]. Nevertheless, without an established theory that explains creativity, it is hard to define the system's creative capabilities.

2.2 Techniques

Since our work is inserted in the field of Computational Creativity, a set of techniques to generate artifacts and create art, either in the form of music, image, text, or problem-solving, is presented in this Section. Papadopoulos, G. and Wiggins, G. [37] have identified some approaches for algorithmic composition which are used in the development of creative systems. In the visual domain, this involves the development of systems that generate images, implementing some techniques described in this Section.

[1] Philosophical discussions of creativity (typified by Boden) usually do not consider the *Typicality* criteria [42].

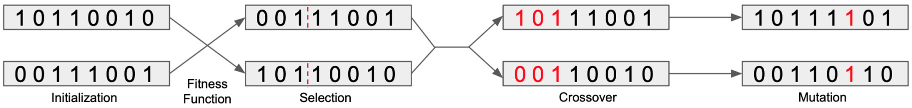


Figure 2.1: Example of a Genetic Algorithm iteration. First, the Initialization, then the Selection and Genetic Operators (Crossover and Mutation). Then, the termination.

In Section 2.4, we present an analysis of the studied systems, as well as their respective implemented techniques.

2.2.1 Evolution

A Genetic Algorithm (GA) is a heuristic search and optimization technique inspired by natural evolution that has been successfully applied to a wide range of real-world problems of significant complexity [29]. It is considered a problem-solving approach in which computers solve problems based on Darwin's theory of the reproduction of the fittest individuals. Analogous to sexual reproduction and natural selection, this happens with the occurrence of genetic operations: crossover and mutation.

There are four steps that a GA performs corresponding to a facet of natural selection [43]: initialization, selection, genetic operators, and termination - figure 2.1. Depending on the chosen approach, when a fitness function is given to evaluate individuals, the higher scored members will survive and proliferate. In contrast, the lowest scored members will die and not contribute to the gene pool of further generations, such as in natural selection. This way, the goal is to find the best combination of elements that maximizes the fitness score, thus accepting a final solution. Each new generation is created through these steps:

1. Initially, an arbitrary number of possible solutions for the problem - individuals - are randomly generated, creating an initial **population** that should hold within a wide range of solutions. This allows for exploring different possibilities over the run of the algorithm.

2. Once a population is created, individuals must be evaluated according to a **fitness function** that takes in the characteristics of a member and outputs a numerical value of how feasible the solution is. Once the fitness values for the whole population are calculated, a specific number of individuals are chosen from the source population to go to the next phase [29].

3. The selected individuals are now used to create the next generation of the algorithm through genetic operators: crossover and mutation. By pairing two individuals, as in sexual reproduction, children are generated with a combination of parents' characteristics. At this step, new genetic material must be introduced into the generation to avoid to become stuck in a local extreme very quickly and not obtain optimal results. This occurs by introducing **crossover** - by randomly choos-

ing a point - and **mutation** - by changing a small portion of the children's genome that they no longer mirror subsets of their parents' genes.

4. Finally, the algorithm must end if it has reached either some fitness threshold of performance, or a maximum number of iterations, returning a final solution.

In turn, when there is human intervention in the run of the algorithm, this model becomes interactive - Interactive Genetic Programming (IGP). In this evolution technique, it is the programmer who decides the best individuals generated from a certain iteration, assigning them a specific value for the fitness evaluation.

Despite having an interesting and successful technique in which the results are satisfactory, this model of programming has some disadvantages [11]:

1. **Speed** - the system needs to stop at every run and wait for a human to assess the generated artifacts.

2. **Coverage** - the system cannot give a human all the possibilities to judge the artifacts, reason why the population is limited.

3. **Human Behaviour** - humans assess some material based on the decision on how it appears at that moment, rather than its evolution on long term potential, which may cause problems in the convergence of the algorithm.

4. **Human Decision** - it diverges the goal of understanding the creative process to the point where people can make a computer genuinely creative.

Nonetheless, some of these disadvantages are also considered when applied to automatic evaluations. For example, the population may be limited since it is impossible for the machine to assess all the possible individuals.

Cartesian Genetic Programming (CGP) is another evolution technique proposed by Miller, J. [32], presented as a general form of genetic programming that represents programs as directed acyclic graphs. Each node in the directed graph represents a particular function encoded by specific a number of genes. The genotype[2] is composed of genes that are integers that represent where a node gets its data, what operations the node performs on the data, and where the output data required by the user is to be obtained. The graphs are represented as a two-dimensional grid of computational nodes that turns the representation simple, flexible, and convenient [32].

Despite that it is hard to find a function that aesthetically assesses artifacts, by taking into account that a GA is a search method that gives a wide range of solutions for a given problem, it satisfies some

[2] The genotype is the part of the genetic makeup of a cell, and therefore of any individual, which determines one of its characteristics (phenotype).

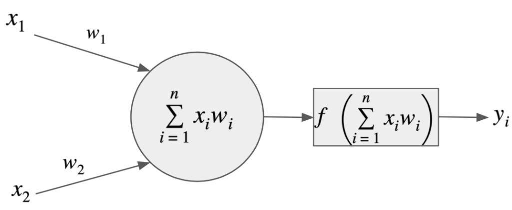


Figure 2.2: Representation of one node and its calculations.

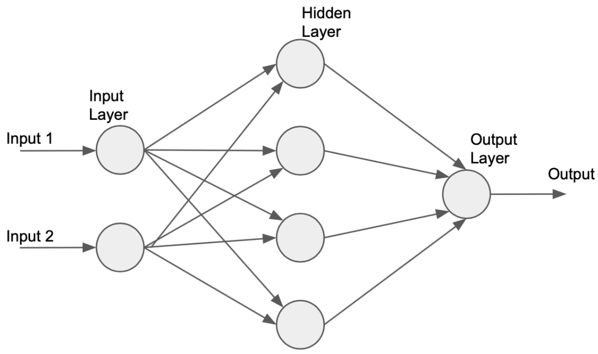


Figure 2.3: Representation of a Neural Network with one hidden layer.

of the requisites a novel and valuable artifact shall have [4]. However, it is a stochastic process and, without human intervention, depends on the fitness function, which may not be the most appropriate for the given problem. Besides, it is not guaranteed that the best solution will be found.

2.2.2 Learning Systems

Artificial Neural Networks (ANN) is a technique that represents a model of how the human brain is structured, thus intended to mimic the way humans learn. The ANN is trained on previous distributions, learning from examples [37].

These networks consist of units of calculation called neurons - nodes - that are connected by synapses - weighted values. This way, given an input, one node will perform some calculation by an activation function - *sigmoid* for instance - multiplying then the result of this calculation by a specific weight as this value goes through the network [23] - figure 2.2. Depending on the size of the network - the number of layers - the weighted result may be the output of the neural network - figure 2.3.

The **back-propagation** [48] is a technique applied to the networks to adjust their hidden layers of nodes, changing the weighted values according to a given learning rate[3], when the output is not what it was expected. In the end, the results may exhibit the statistics of the training set.

Generative Adversarial Networks (GAN) is another model introduced by Goodfellow et al. [20] in

[3]The learning rate controls how much the model changes each time the weights are updated, in response to the estimated error.

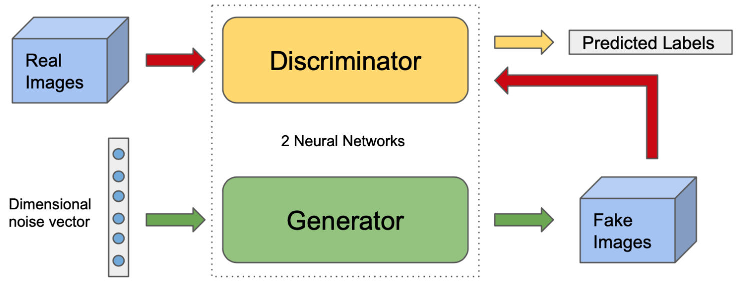


Figure 2.4: Block Diagram of a Generative Adversarial Network.

2014, that generates new artifacts that exhibit the statistics of the training set. The difference from the ANN is that this model is composed of two neural networks - the generator and the discriminator. While the former generates new data instances that look authentic, the later decides whether each instance of data produced belongs or not to the used training set. The goal of the generator is to fool the discriminator about the authenticity of the artifacts produced. In contrast, the goal of the discriminator is to identify fake artifacts produced by the generator.

The steps taken are the following - figure 2.4:

- The generator takes random numbers as input, returning an image.
- The discriminator takes the image alongside a stream of images from the training set.
- The discriminator predicts the authenticity of both fake and real images, returning then a probability between 0 (fake images) and 1 (real images).

As an example, training a GAN on photographs may lead to the generation of pictures that look authentic to human observers, having such realistic characteristics [20].

Although the artifacts produced by a Neural Network look interesting and aesthetically pleasing, they are, by definition, unable to produce novelty as humans do:

- To produce music, they need to be trained in a huge amount of previous melodies that which structure will be reproduced by the networks when composing the output.
- To produce or generate images, the networks learn to reproduce statistics of previous sets of images.

Similarly, the generation of poems or texts requires these networks to be trained to reproduce a given structure that already existed in previous data. Besides, it requires a filtering process of the training set to avoid conflicts, which may take a long time, depending on the problem.

2.2.3 Niche Construction

According to McCormack, J. [30], Evolutionary Music and Art is ”to devise unique kinds of evolutionary software instruments that offer the possibility of deep creative exploration of generative computational phase-spaces”. New approaches in the field of creative systems have emerged to design specific components and their interactions instead of using aesthetic fitness evaluation. Their features are the following [31]:

- An ecosystem consists of components with carefully designed interactions between themselves and their environment.
- The ecosystem operates and is conceptualized within the generative medium itself. For example, a sonic ecosystem operates in the medium of sound, rather than being a ”sonification” of some other process.
- Components within an ecosystem are interconnected in such a way that they can modify their environment, often to their benefit or that of their descendants.

Niche Construction is the process whereby organisms modify their own and each others' niches by considering that are resemblances between characteristics of creative ecosystems and real ones, such as heterogeneity, diversity, mutualism, stability under change, and complex feedback loops [19].

Considering a system composed of drawing agents that move around over a two-dimensional surface, leaving an ink trail as they go, McCormack, J. and Brown, O. [31] proposed to add to these agents a new allele - local density preference - in its genome. This new feature defines the preference for the density of lines already drawn on the canvas in the immediate area of its current position, taking into account the ratio of black canvas per unit area. Without it on the genome, agents tend to move homogeneously on a white canvas. The characteristics of the path the agent decided to go through are determined by a stochastic process, having the following alleles on its genome [31]:

- **Curvature** - from straight line to a maximum curvature rate.
- **Irrationality** - controls the rate and degree of change in the rate of curvature (figure 2.5).
- **Fecundity** - probability of reproduction
- **Mortality** - probability of dying

However, with the new allele, by adding a niche construction process in creative contexts, the drawing that previously tended to be homogeneous, in terms of the style and overall tonal density observed, becomes heterogeneous and diverse. It also exhibits significant aesthetic variation without human intervention: agents receive discrete fitness rewards if each of the local environmental conditions is within their preferred range. Besides, depending on the path, some agents obtain greater fitness rewards.

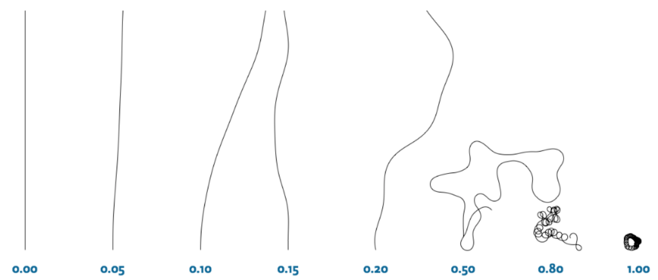


Figure 2.5: Different measures of irrationality applied to line drawing.

Similarly to this, other techniques to do paintings or drawings have been explored. For example, Moura, L. [34] started in 2001 the first version of a system composed of little car robots that have the ability of line drawing with colors (red, green, blue) in a white canvas, leaving an ink trail as they go. For sixteen years, the system is being improved to perceive movement and nearby activity through sensors that influence the robot's path and activity. A kind of negative feedback determines the finalization of the robots' activity - robots stop reacting because a specific density of color is achieved [34]. According to its creator, "the idea here is to show that the humans are not the creators of the final product but rather the creators of the process that built that job". As any artist, these robots are stimulated by everything that surrounds them, either the environment, sound, or interactivity with people.

2.2.4 Other Techniques

Papadopoulos, G. and Wiggins G. further present other AI techniques for algorithmic composition, as well as their advantages and disadvantages [37]:

- **Mathematical Models** - Probabilistic models that allow to represent randomly changing systems - these models are, for instance, stochastic processes or Markov chains. Regarding the musical domain, it composes pieces in different genres, deducing the rhythm of the melody using Markov chains, choosing then the pitches at a later stage. Although these stochastic processes help on the understanding of the domain in analysis, they have some disadvantages [37]. Firstly, it is necessary to analyze many pieces to find probabilities for style simulation. Secondly, it is important the deviations from the norm, and how it incorporates the music. An example of these systems is the Cybernetic Composes [1], a system that composes music in different genres (jazz, rock, or ragtime) [37].
- **Knowledge Based Systems** - Systems that make use of rules and constraints to explain the choice of actions for a given complex problem. The main advantage is that it is constructed to reason for a specific domain - thus, the knowledge given to the system is thorough for that domain. Even though most AI systems for algorithmic composition are knowledge-based systems,

there are disadvantages [37]. Firstly, the time consumption for knowledge elicitation. Secondly, the complexity introduced by the ”exceptions to the rule”, and their preconditions (necessary in the domain). Moreover, since these systems do what it is programmed with the guidance of the ”expert”, they hardly depend on it. An example of these systems is the CHORAL [12], a rule-based expert system for the harmonization in the style of J.S. Bach, which creates playable improvisations from a set of potential actions as an initial state for the musical problem.

- **Grammars** - Based on the usage of production rules for strings, which have no semantic purpose since they only describe how to create valid strings with the alphabet letters. By using these grammars, it is possible to define any languages, from natural to programming ones. In creative purposes, grammars are defined with rules that create creative artifacts. Regarding musical composition, Papadopoulos, G. and Wiggins G. [37] identify some disadvantages. Firstly, most musical grammars do not make strong claims about the semantics of the pieces. Secondly, parsing is computationally expensive, especially if we are working with ambiguity.

  An example of grammars systems is the Experiments in Musical Intelligence (EMI) [10], a project focused on the understanding of musical style and stylistic replication of various composers to extract signatures using pattern matching.

In the end, some of the systems can implement more than one technique: EMI, for example, is categorized as a grammar, but it can also be seen as a knowledge-based system.

2.3 Systems

In this Section, we describe some systems that implement techniques presented in Section 2.2. Our goal is to generate abstract images inspired from music, therefore we focused our analysis mainly in the musical and visual domains. The analysis in the music domain gives us knowledge about music features that can be mapped. In contrast, the analysis of the visual domain allows to understand what are the mechanisms used to generate images.

2.3.1 Music2Art

The generative system Music2Art was developed by *Artrendex Inc*, a New York-based startup that builds innovative AI for the creative domain [35]. It takes music as inspiration to make art by learning aesthetics from a collection of images and videos. The system, in real-time, takes the music stream from a concert, decomposes it to basic frequencies, and aligns the audio frequency contents with a representation of the aesthetics previously learned.

This system was first implemented with a model of Creative Adversarial Networks (CAN)[4] that generates art with creative characteristics, by looking at art and learning about style. It can continuously learn from art that was already generated, adapting its generation based on what it has already learned. The artifacts produced with the CAN model simulate a given distribution from the training set, generating images that are novel, aesthetically appealing, and that do not emulate the art distribution [14].

To the development of the Music2Art system, two models were proposed:

1. **Art-Generating Agent**: the agent's goal is to explore the creative space by deviating from the established style norms, generating novel art with increased levels of "arousal potential". The agent has a memory that encodes the art it has been exposed to, learning it, and updating that memory with the addition of "new art". This way, by increasing the stylistic ambiguity and deviations from style norms, the agent tries to avoid moving too far for what is commonly accepted as art.

2. **Functioning Model**: this model was designed to generate art that maximally confuses human viewers to which style it belongs to, not following established art movements or styles. It complements the previous one - the Art-Generating Agent.

Despite the success these systems have to generate images, the GAN model is limited in their ability to generate creative products in their original design. This way, modifications were proposed to its goal to make it capable of generating creative art by maximizing deviation from established styles, and minimizing deviation from art distribution, while staying within it [14].

Although both models have the same architecture, the CAN distinguishes from GAN because a new signal from the discriminator is added, achieving the three following goals:

- Generate novel work.
- The novel work is not too novel.
- The generated work increases the stylistic ambiguity.

Regarding the CAN model, the first signal is a discriminators classification of "art or not art". In turn, the second is a signal of how well the discriminator can classify the generated art into established styles. These two signals are contradictory forces since the former pushes the generator to generate work that the discriminator accepts as "art", while the latter penalizes the generator for doing that: it pushes the generator to create style-ambiguous work. Nevertheless, these two signals together push the generator to explore parts of the creative space that are close to the distribution of art. Besides, the ambiguity of the generated art is maximized, concerning how it fits on the learned labeled art styles. In other words, a new force is added to push the generator to explore the creative space. A visual representation of this model can be seen in figure 2.6.

[4] This AI process of emulating art with Creative Adversarial Networks - Elgammal et al. [14] - is built over Generative Adversarial Networks - Section 2.2.2.

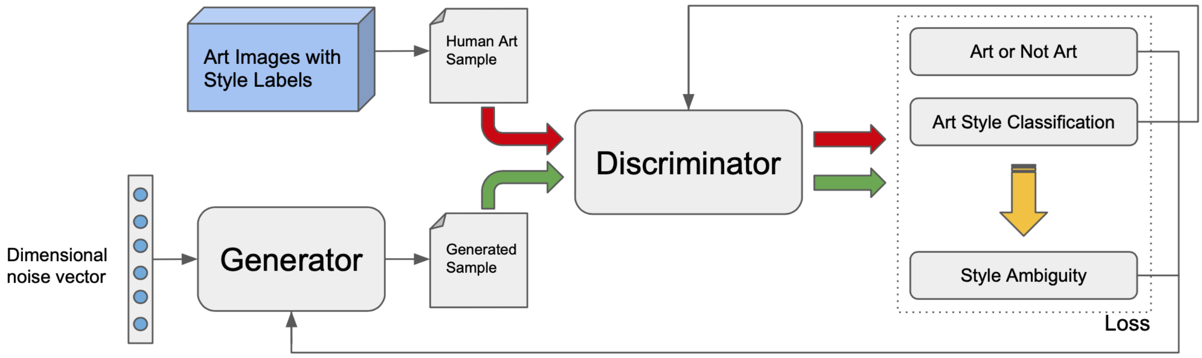


Figure 2.6: Diagram of Creative Adversarial Networks. The Networks (Generator and Discriminator) work together to explore the creative space, generating novel images that are classified as "art" by optimizing a criterion that maximizes stylistic ambiguity while staying within the art distribution.

Finally, the Music2Art system can render new images that are directly reactive and synchronized with the music at a live concert, giving to the audience a pleasing multi-sensory experience [35].

2.3.2 Evolutionary Art

The Evolutionary Art system was built by DiPaola, S. and Gabora, L. [11] to evolve abstract portraits of Darwin using an automatic fitness function that rewards not only genomes with a likeness of Darwin, but also genomes that exhibit specific characteristics of human artists techniques. It incorporates the CGP, the genetic programming technique that, for the same genetic operators (crossover and mutation), differs in how the program is represented: a directed graph of indexed nodes [32].

The system considers that each node of the graph has an input (x and y coordinates of a pixel), and a function that outputs the color's hue, saturation, and value for that pixel. According to the author [11], one individual in the population is manifested as one program that runs successively for every pixel in the output image, then tested against the fitness function. The functions used for each pixel in each node take into account reactions to the **color** and **position** of the model image so that it can better approximate human portrait drawing techniques.

The algorithm absorbs the existence of a limitation of traditional algorithmic art, in which the program relies on human intervention, where the creator selects, in each iteration, the most aesthetically pleasing variants of a generation to produce the next one [11]. Based on this, the algorithm evolves art without human intervention, employing an automatic **fitness function** specific for portrait painting.

The fitness function presented has to do with face versus background composition, tonal similarity over exact color similarity, and unequal dominant and subdominant tone. Thus, it calculates four separated scores: one for resemblance, and the remaining for rules presented above. Consequently, it fluidly combines the four scores in various ways to mimic human creativity, trying not to remake the Darwin portrait, but to explore a family tree of related portraits that inherit creative painting strategies through an evolutionary process. Once an aesthetically pleasing portrait is created, the program can use the same painting strategies on any new model image [11].

2.3.3 StackGAN

The Stacked Generative Adversarial Networks is a system developed by Zhang et al. [49] that synthesizes photo-realistic images from text descriptions, by decomposing the main problem (initial textual input) into more ageable sub-problems through a sketch-refinement process.

The generation of high-resolution images with photo-realistic details is decomposed in two-stages:

- **Stage-I** GAN model sketches the primitive correct colors and rough shape based on the given textual description, and draws the background layout, providing low-resolution images.
- **Stage-II** GAN model takes the results from the previous step to correct them, generating photo-realistic details and diversity, based on the text descriptions as input that are omitted by the previous stage.

These two models are alternatively trained to compete with each other. The generator is trained to reproduce true data distribution, and the discriminator is optimized to distinguish generated images from real ones. In other words, the generator tries to produce photo-realistic images that fool the discriminator as being real ones [43]. Due to the limited number of training text-images pairs (that makes the training of the GAN a challenging issue - Section 2.2.2), an additional technique was introduced - Conditioning Augmentation - to allow the generator to generate images conditioned on some variable from an independent Gaussian distribution, where the mean and diagonal variance matrix are functions of the text embedding. By using this, the diversity of the synthesized images improves, and the training of the network stabilizes [49].

The results demonstrate the capability of the proposed system in comparison with other generative text-to-image models, by generating higher resolution images with more photo-realistic details [49]. Moreover, it takes into account that images can fail to contain necessary details, the reason why a sketch-refinement process was introduced in which the generation is decomposed in two-stages. However, it once again generates images from a previous set of images used in training, reproducing its statistics and structure.

2.3.4 The Painting Fool

The Painting Fool, developed by Colton, S., is a computer program and an aspiring painter that has decision-making abilities in the field we are working on - Computational Creativity [6]. Its goal is to produce increasingly interesting and culturally valuable artifacts that are considered art, such as paintings, sketches, and drawings. This includes new ways to construct paintings in terms of scene composition, choice of materials, or painting styles. It uses evolutionary search and photo-realistic rendering abilities, engineered and further trained to transcend most generative art projects by addressing a higher level of artistic behavior, such as critical ability and cultural awareness [8]. The program makes use of pipelined AI techniques to generate scene elements (according to a user-defined evaluation), a visual analysis system to evaluate features on an image (for instance, symmetry, texture or color variance), and processes that segment images into a set of color regions. It creates novel paintings, either by user-defined input, such as photographs, or by applying knowledge learned from visual grammars.

It first starts by segmenting the entire image in several paint regions, and abstracts its borders to be reasonably smooth. Then, it takes each shape in each segmentation region and renders it with simulated art materials, such as acrylic paints, pastels, or pencils, producing an artistic rendering of the image. Besides, the software was initially trained with a knowledge base of different settings for the segmenting and rendering process. Its behaviors are based on simulations of the physical painting process, by looking at digital photographs and determine regions of colors [6]. Then, it abstracts these regions and changes their color according to palettes, rendering them stroke-by-stroke, simulating the natural media, as well as their usage on outlining and filling paint regions [8]. The structural elements are themselves generated by context-free grammars, while the abstract art pieces and image filters are generated with evolutionary methods. In the end, it uses Machine Learning to predict when two abstract images have too much structural similarity.

The author [8] considers that this software, to be called a painter, must exhibit both cognitive and physical behaviors common to human painters, being trained to act increasingly more creative. Even though the system can work as an expert, it needs to be told the emotion expressed in the image to decide from its knowledge base how it would render the painting, which is a slow and non-autonomous process.

2.3.5 ETC Video Synthesizer

The ETC video synthesizer, produced by Critter & Guitari [18], is a system that uses the idea of creativity as a way to explore computers and computation, programming more generally. According to its creators, it is a self-contained device that takes an audio input and then runs ”creative programs written in Python” that use audio coming in trigger events to adjust their visuals and adapt to accompany it. While playing music, by having a projector or a big screen, this synthesizer will be generating visuals and shooting them out, taking in the audio running. The synthesizer knobs are used to parameterize values written in the code to adjust the visual output colors, shape, and size in real-time. In contrast, the scope mode with audio input waveform is used to trigger events on the generated output, if some amplitude is reached.

The ETC can load different programs known as ”Modes” that can be geometric, dynamic, text, or image-based. Two types of ”Modes” can run together:

- **Scope Modes (S)** that continuously track the amplitude of incoming audio to generate an image.
- **Trigger Modes (T)** that are rhythmic in nature, and change imagery when the audio level crosses a certain threshold.

This system generates art having sound as input. However, since it creates video animation and not static images, it always needs a synthesizer to accomplish its work [18].

2.3.6 Visual Information Vases

The Visual Information Vases (VIV) is an AI-based generative art system proposed by Horn, B. et al. [22] that focuses on the evocation of inspiration from a source domain to create an artifact in a different domain through cross-domain analogy mappings. The system uses as model of inspiration to produce 3D-printable vases 2D images uploaded by a user. It attempts to create a vase with similar aesthetic measures to those of the inspiring image through evolution. Results are diverse and functional creations.

VIV analyzes the colors of a user's image to create a color palette from salient and dominant colors, creating an aesthetic profile for the image. Then, an evolutionary algorithm is used to produce a vase with a similar profile to that supplied by the user. By evaluating the image's color palette, the algorithm scores the image along with four aesthetic measures: activity, warmth, weight, and hardness. The fitness function used is the Euclidean distance between the aesthetic profile of the image and the generated vase's profile [22]. The resulting 3D vases can be printable to produce a functional and decorative vase.

Even though the system can create diverse vases that reflect the aesthetic profile of the inspiring images, thus evoking similar feelings, it uses a very abstract and limited representation of the image. It only considers colors to generate static and simple vases, which may limit the results. Besides, the system does not consider how colors in the image relate to each other, so the output vases have no color or patterns that could reflect some essence of the inspirational image. Nonetheless, this system implements a novel cross-domain inspiration framework that translates aesthetic qualities from colors to vases, resembling methods used by human artists to create content with external inspiration sources. It may prove that by using inspiration, generative systems can create artifacts from a wide variety of sources [22].

2.3.7 Cross-Domain Analogy: From Image to Music

In 2017, Teixeira, P. and Pinto, H. [47] described a system that proposes an association between the visual and musical domains by generating music from images. This inspirational system generates musical artifacts given an image as input, by identifying a set of features to be extracted from an image.

Therefore, by considering all the extracted visual features, the system interprets and uses them as a starting point to translate into several components of the musical domain. It generates three different outputs, whose results are promising, since the majority of listeners classifies both versions as music [47]:

- **Raw Version** - pure translation of the visual features into the musical features is attempted.
- **Harmonized Version** - visual features are processed, taking into account the whole image. Some musical conventions are imposed to create a more aesthetically pleasurable musical artifact.
- **Genetic Version** - it uses the two previous versions to generate the initial population of the GA. This version attempts to improve the aesthetic value of the artifacts by mainly considering the similarity to a Rock-type progression.

The system uses information extracted from the colors of the image - temperature, variations, groups of the same color - and influences the final music features - pitch, octave, volume and duration of each note, the Beats Per Minute (BPM) of the music, the generation of a chord sequence and percussion [46].

By exploring this analogy, the system aims at generating musical artifacts that could reflect the images that they were inspired. It outputs the three different versions using one possible approach between the two domains, and genetic algorithms to generate music with improved aesthetic value.

2.4 Analysis of the studied Systems

Table 2.1 summarizes each system described in Section 2.3, displaying information about the domain and techniques - Section 2.2 - used to achieve its goal.

The explored related work gives us information and mechanisms that will be considered when developing our system. Since the goal is to generate visual artifacts, we are focusing our attention in the visual domain, the reason why we further studied systems that use techniques that generate images. Thus, to build a cross-domain association, we must understand which features of each domain relate to each other, and how to connect them.

As presented in Section 2.2, there are different techniques we can follow to generate creative images. For instance, we can use Machine Learning models to learn from existing artifacts and create new, similar ones, or we can use Evolutionary techniques to combine and ”shuffle” existing work, providing a way to search for better results while maintaining certain randomness. However, different techniques have both advantages and disadvantages, depending on the goals of the system. First, if one is using Machine Learning models, the system usually generates images that follow features that already existed in previous work, which may lead to the creation of artifacts that have no novelty. Second, if the system uses Evolution techniques, the fitness functions in use may not be the most accurate to the goal we are

System	Domains	What	Techniques
Music2Art	Musical and Visual	Generates synchronized art with music in real-time by learning about styles and deviating from style norms	Creative Neural Networks
Evolutionary Art	Visual	Evolve abstract portraits of Darwin using an Automatic Fitness Function	Cartesian Genetic Programming
StackGAN	Textual and Visual	Generates photo-realistic images conditioned on text-descriptions	Generative Neural Networks
The Painting Fool	Visual	Generates artistic images in various different styles	Context-Free Grammars, Machine Learning and Evolutionary Methods
ETC Video Synthesizer	Musical and Visual	Takes an audio input and then runs creative programs written in Python to generate video	Direct Translation
Visual Information Vases	Visual	Generates 3D vases having inspiration from images	Genetic Algorithms
Cross-Domain Analogy: From Image to Music	Musical and Visual	Proposes a mapping between the visual and musical domains to generate music from images	Direct Translation and Genetic Algorithms

Table 2.1: Overview of the studied systems

looking for, or the execution time may be too long, especially if the fitness evaluation in each iteration is interactive and performed by humans - as explained in Section 2.2.1.

In the end, despite all the disadvantages they have, the studied systems give us a background on how to generate visual artifacts that are aesthetically pleasing, either abstract or photo-realistic images, paintings, sketches, or drawings, having as inspiration artifacts from other domains, such as visual, textual or musical domains.

3

Musical Domain

Contents

3.1 Definitions

Music is considered a form of art that usually combines sounds, following arrangements over time - it has a duration and a well-defined starting and ending points. According to the dictionary[1], it is defined as *"an art of sound in time that expresses ideas and emotions in significant forms through the elements of rhythm, melody, harmony, and color"*. It is a human practice that varies from culture to culture, and it can be considered creation, representation, or communication. In Western cultures, when analyzing music, one follows the fundamentals of music theory and concept rules that create the formal language used by composers.

Musical composition is usually divided into three main parts: **harmony**, **melody** and **rhythm**. Music can be considered as an expression of harmonic frequencies, melodic events and rhythmic values. The way these are organized depend and vary from culture to culture, but these three elements are commonly used in Western music. For our work, we considered Classical music pieces as our starting point. Regarding the music's structure, we consider that symphonies are divided in movements, that are subdivided in sections, and subdivided in measures, which are subdivided in individual **notes** - the minimal unit of music - and **chords** - three or more notes played together. While a note may last less than a second, a symphony may last more than an hour. It is important to note that this division is relative and that there are elements that rule each of these parts. In the following sections, we further explain some of the concepts in which music is divided. It is important to note that what is explained below comes from the author's knowledge acquired throughout this year from other colleagues [46] [5] [44], as well as from music theory sources available online [39].

3.2 Parts or Voices

Parts are generally represented by instruments (or voices) and refer to a single strand, or melody or harmony of music, within a larger ensemble musical composition. When analyzing a score, each line of the sheet represents one part of the music, that is, one instrument (or voice). Each score may have different parts, that may or may not contain the same instruments, arranged to sound together. In turn, each instrument is characterized by its sound quality - the timbre of the produced sounds. According to its definition, *"timbre is that attribute of auditory sensation in terms of which a listener can judge that two sounds, similarly presented, and having the same loudness and pitch, are dissimilar"* [41]. Therefore, timbre is the perceptual attribute that enables humans to distinguish among sounds that are playing the same tones, equally loud. It is important to note that two instruments can play the same note, having different timbres.

[1] https://www.dictionary.com/browse/music

Tempo	Description	Beats per Minute
Largo	Slow and solemn	40 - 66
Adagio	Slow (literally ”at ease”)	66 - 76
Andante	Moderate walking speed	76 - 108
Moderato	Moderate speed	108 - 120
Allegro	Fast	120 -168
Presto	Very fast	168 - 208

Table 3.1: *Tempo*, its description and BPM.

3.3 Measures

In Western Music, each staff[2] is usually divided by vertical lines. This division is called a measure (or bar) and, when analyzing a musical sheet, it provides a way of quantitatively divide the sounds of a musical composition into groups, based on its beats. The boundaries of each staff provide regular reference points to pinpoint locations within a musical composition, making written music easier to follow. Each measure contains different elements - for our purpose, we are going to call them events - that can be *Tempo* (Section 3.4), Time Signatures (Section 3.5), Notes (Section 3.6) and Chords (Section 3.8).

3.4 *Tempo*

The ***tempo*** indicates the speed or pace of a given music (or music subsection). In classical music, it is usually represented with an instruction at the start of a piece (often using conventional Italian terms), although it can be represented by a numerical value measured in BPM - table 3.1. In practical terms, a *tempo* of 60 beats per minute can be indicated by the instruction ”Largo”, and means one beat per second, while a *tempo* of 120 beats per minute - ”Moderato” - is twice as rapid, that is, one beat every 0.5 seconds.

3.5 Time Signatures

Throughout the music, each measure has the same duration given by the time signature - that is, how many beats each measure contains. It is typically at the beginning of the staff, represented by two values written as a fraction. The top number is the number of beats to count - the number of beats a measure has. The bottom number is the note value, that is, the kind of note to count - quarter notes, eighth notes, or sixteenth notes - figure 3.1. For example, a time signature of $\frac{4}{4}$ indicates that the beat unit is a quarter note, and each measure has four of these beats. Both the *tempo* (Section 3.4) and time signature characterize the **rhythm** of the music.

[2] Set of five horizontal lines found on music sheets.

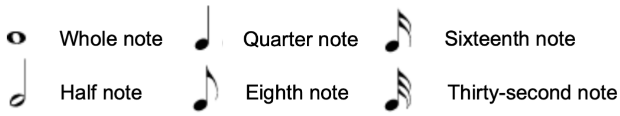


Figure 3.1: Note duration.

3.6 Notes

Notes are the minimal unit that music contains, and represent sounds that are formed by one mode of vibration of the air. Each note has a duration and pitch, the two essential concepts a written piece of music needs to define a note. The former describes how long, temporally, the note lasts, while the latter is related to how high or low one sound is in comparison with another.

There are different note lengths according to the amount of time the note sounds - its **duration**. All the note lengths are defined by how long they last compared to the whole note - the simplest-looking note - figure 3.1. All other note lengths are defined by how long they last compared to the whole note. A note that lasts half as long as a whole note is a half note. A note that lasts a quarter as long as a whole note is a quarter note. The pattern continues with eighth notes, sixteenth notes, thirty-second notes, and so forth, each type of note being half the length of the previous type. The note length works like fractions: two half notes will be equal to (last as long as) one whole note.

Pitch is related to the **frequency** of one sound produced by a note, describing how high or low one sound is. These two concepts - pitch and frequency - are related but express different ideas. Pitch is subjective and depends on the perception people have of the sound, while frequency is objective and represented by numeric values, measured in Hertz. When writing a musical sheet, composers relate the notes to one musical alphabet, building a clear and comprehensive method of music manifestation. This association makes it possible to match the frequency to common words known by musicians. The interval between two pitches is measured in tones[3], where the smallest is the semitone[4]. In Western music there are twelve ordered pitch-classes[5], designated by the letters of the alphabet, from A to G. These pitch-classes can be represented in a circular diagram, the Chromatic Circle - figure 3.3.

The difference between two notes is the **interval**, and its measure unit is the tone. If a third note is added to it, we lead to the construction of one chord - the conjunction of three or more notes that are most of the times sounded simultaneously.

There may also exist one symbol attached to each note, representing one semitone higher or lower than the desired note. This symbol - the accidental - may be a flat ($\flat$), that translates the note into one semitone below, or a sharp ($\sharp$) that translates the note into a semitone above. For example, a note represented by a letter followed by a flat symbol (C$\flat$) is a halftone lower than the pitch letter of the note

[3] In musical notation, tone - or whole step - is used when the distance between two consecutive notes is a whole note.
[4] In musical notation, semitone - or half step - is used when the difference between two consecutive notes is half a tone.
[5] In musical notation, a pitch-class is the set of all pitches that are a whole number of octaves apart.

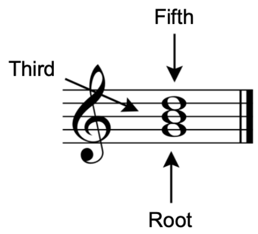


Figure 3.2: Chord triad.

(C), while a note represented by a letter followed by a sharp symbol (C♯) is a halftone higher than the pitch.

Each note may also be represented by its pitch letter followed by a number - the **octave** in which the note is. One octave is the interval between two musical pitches, in which one has double frequency than the other. For example, the middle C in the piano that corresponds to the fourth octave (C4) represents the pitch C in the fourth octave. One octave above has double frequency, and it is represented by C5, the C pitch in the fifth octave.

Notes can be played with different intensities, indicated by the **volume** - it represents the variance between a weak and a strong sound, according to its loudness.

3.7 Scales

In Western music, the scale is any set of musical notes that can be ascending or descending from the twelve pitch-class, according to its steps. For example, the major scale component pitches are C - D - E - F - G - A - B, with a well-defined sequence of steps: tone, tone, semitone, tone, tone, tone, semitone, repeating the cycle. On this scale, we start with the C note and follow a well-defined sequence of intervals until the return to the do note again, setting up the major scale. There are other scale types, but for our work we considered the **Chromatic Scale** - that includes all the pitches in a single octave - twelve pitches with a distance between them of a semitone (C, C♯, D, D♯, E, F, F♯, G, G♯, A, A♯, and B) - figure 3.3.

In music theory, a scale may have a tonal center, a key, which is usually the first degree of a scale. This way, the music's scale, or the note sequence that is used to compose melodies and harmonies in the key of the tonic, is built around this tonal center. The key is the so-called tonic, and its scale may vary from music to music, or from style to style.

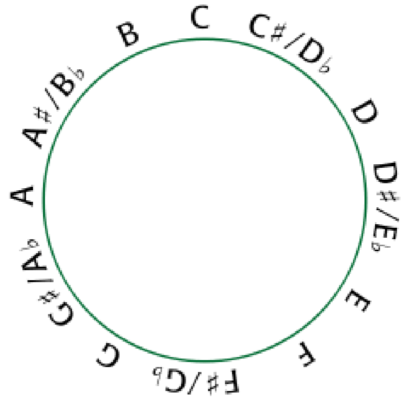


Figure 3.3: Chromatic Circle.

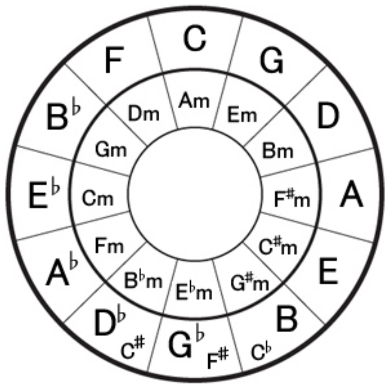


Figure 3.4: Circle of Fifths.

3.8 Chords

A **chord** is a combination of three or more notes played simultaneously, defined by their root note and their quality (major or minor, for example). They are usually triads, that is, a chord made of three notes that consists of a root - the lowest note - and two other above it - the third and fifth - figure 3.2. Triads are named according to their root, and the number of steps between it and the other two notes - the distance above the root. In Western music, the most common chords are major, minor, diminished and augmented, although for our work, we only consider the first two:

- **Major** - The third is four semitones above the root, while the fifth is seven semitones above the root. For example, the notes of a C major chord are the first (the root note), third, and fifth notes, which are C (the root note), E and G [46].
- **Minor** - The third is three semitones above the root, and the fifth is seven semitones above the root. For example, C, E♭, G [46].

In music's theory, the **Circle of Fifths** - figure 3.4 - is a circular diagram that represents the relationship among the twelve pitch-classes of the **Chromatic Scale**, their corresponding key signatures, and the associated major and minor keys, a consequence of the foundations of Western music, where a lot of musical knowledge is concentrated [5]. This way, it shows how all the musical notes, keys, and chords relate to each other. Major keys all have a relative minor key - for instance, the relative minor key for G major is E minor. The relative minor for D major is B minor. A major key and its relative minor key make use of the same scale. Thus, this circle is used to help when writing music, on predicting chords in a chord progression, and on understanding relationships between major and minor keys.

4

Visual Domain

Contents

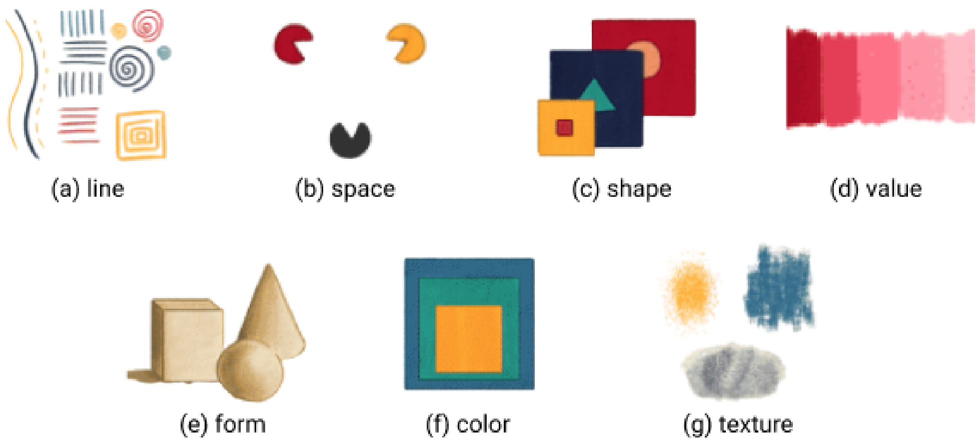


Figure 4.1: 7 Elements of Art.

Figure 4.2: Tone's Value Scale.

4.1 Definitions

For this Thesis, we targeted images that do not reflect or convey anything "concrete" or "real" of the real world. When looking at an image, there may exist details that one can identify, either in the background[1] or foreground, such as lines, colors, textures, shapes, spaces, among others.

Image interpretation is the act of examining images to identify its objects and analyze their meaning, involving a considerable amount of subjective judgments dependent on the capability of generalizing at different levels of complexity. The machine's ability to analyze and process images to retrieve their features can be compared to the perception humans have about one image and their elements. As well as with music pieces, it is noted that this interpretation is personal and may vary from culture to culture.

According to [15], there are Seven Elements of Art, the "building blocks" to create images on canvas - line, space, shape, value, form, color and texture - as depicted in figure 4.1. When generating images, paintings, or drawings to compose a visual piece of art, these elements may be manipulated and mixed in with design principles. Every work does not need to contain all of these elements, but usually at least two, no matter which ones, are always present [15]. In the following sections, we further describe each of the previous elements, especially **color** and **shape**.

[1] According to the dictionary, the background of an artistic image is *"the part of a painting representing what lies behind objects in the foreground - the closest view plan for the viewer ".*

Figure 4.3: Color Wheel.

4.2 Color

Color is one of the main elements used in visual arts, often considered the whole point for people who are visual learners and thinkers [17]. Symbolically, colors can have different meanings and transmit different perceptions and sensations depending on what is being analyzed, the time of the history, culture or personal beliefs. In Western cultures, for example, colors are often related to the emotions of human beings, thus conveying a mood. Conventionally, they may have more than one emotional interpretation, expressing certain events or sensations, transmitting different ideas, as in the case of traffic lights, where green, yellow and red refer to different actions.

Colors are computationally defined by three measurable attributes: hue, value, and intensity, as previously mentioned. Hue stands for the color itself, value for the brightness of the hue, and intensity for the quality that distinguishes a strong color from a weak one. According to [33], there are three basic categories used to create a logical structure of color - the **Color Wheel** - figure 4.3, the **Color Harmony**, and the **Context** of how colors are used.

The first circular diagram of colors was developed by Isaac Newton in 1666 [33] and since then, artists and scientists have studied and designed numerous variations of this concept. This way, the Color Wheel is a diagram that organizes color hues around a circle, showing the different relationship between primary, secondary and tertiary color, based on red, yellow and blue hues, traditionally used in the field of art:

- **Primary Colors** - Red, yellow and blue. These three hues can not be mixed or formed by any combination of other colors.

- **Secondary Colors** - Green, orange and purple. These hues are formed by the combination of any two hues of the primary colors.

Figure 4.4: Different Color Contexts.

- **Tertiary Colors** - Yellow-orange, red-orange, red-purple, blue-purple, blue-green and yellow-green. These hues are formed by mixing equal parts of primary and secondary colors.

Regarding the other two categories used to create a logical structure of color - Color Harmony and Context - and considering what is described in Section 3, harmony is defined as a pleasing arrangement of elements, whether in music, color, or any other domain. In the visual domain, harmony is pleasing to the eye, engaging the viewer and creating an inner sense of order, a balance in the visual experience, delivering visual interest and a sense of order - Section 4.2.2. How color behaves with other colors and shapes is a complex area of color theory [33] that can be observed in figure 4.4, where for the same red square, it is possible to perceive the contrast effects of different color backgrounds. Red appears brighter against a black background and duller against a white one. In contrast with orange, red seems lifeless, and in contrast with blue-green, it exhibits brilliance. Besides, the red square appears larger on the black background than on the others. Therefore, depending on the context, the relationship of values, saturation, and luminosity of respective hues can be perceived differently.

4.2.1 Color Models

A color model is an abstract mathematical model that describes the way colors can be represented as tuples of numbers, typically as three or four values of color components[2]. When a color model is associated with a precise description of how the components are to be interpreted, the resulting set is called “color space” - the whole range of colors that a specific type of color model produces. In this Thesis, we studied three different color models to understand how each color can be represented with its different properties [25]:

- **RGB** - stands for red, green and blue hues, used as the three primary colors which span a 3D space - a cube where each component is a Cartesian coordinate - figure 4.5. To generate a wide range of colors, it uses light to create the color, and the intensity of these primaries are calibrated to range from 0 to 1, where color is described by specifying the intensity levels of red, green and blue. In this situation, the color’s ”value” refers to the strength of the colors in relation to each other. It is an additive color model - when the three colors of light are shown in the same intensity at the same time, they produce white, while when they are all out, it produces black.

[2]The number of color components may vary according to the color model used.

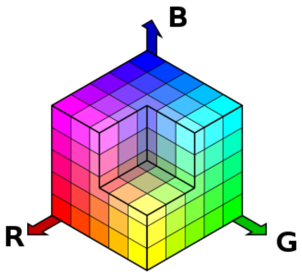


Figure 4.5: The RGB 3D space.

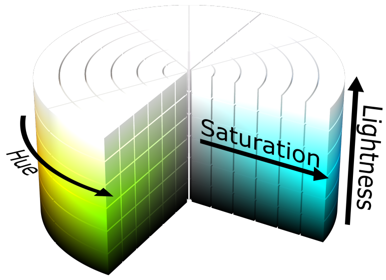


Figure 4.6: The HSL Color Model.

- **HSL** - stands for Hue, Saturation and Lightness. This model is an alternative representation of the RGB color model, where each hue is arranged in a radial slice, around a central axis of neutral colors which ranges from black at the bottom to white at the top - figure 4.6. It uses the three main characteristics of colors - hue, value (saturation) and intensity (lightness) - and its components to describe colors. In this context, lightness is the amount of black or white in one color - increasing lightness adds white to the hue while decreasing lightness adds black. On the other hand, saturation in the HLS model is a measure of the "purity" of a hue - as saturation is decreased, the hue becomes grayer.

Regarding these two color models, a fourth component can be added to represent the **alpha channel**, defining the color's opacity.

Finally, the last Color Model we have studied is the CIELAB (*Commission Internationale d'Eclairage* LAB), that expresses color as three values, **L***, **a*** and **b*** - while L* stands for lightness, a* stands for Red/Green value and b* for Blue/Yellow value - figure 4.7. It was designed so that the same amount of numerical change in their parameters correspond to the same amount of visually perceived change. Thus, this model is intended to be perceptually uniform, meaning that the space between mapped colors corresponds to their visual differences. Based on this, a mathematical formula can be used to quantify the amount of color variation - **Delta E**. The output indicates the status of a color according to a reference, that is, the difference (or distance) between two colors - the closest to 0, the smaller the difference between them.

There are other Color Models, and no one is "better" than another, since they all can represent a wide range of colors. Typically, the choice for a color model is related to external factors, such as a graphics tool or the need to specify colors with specific and detailed characteristics.

4.2.2 Chromatic Harmony

Chromatic Harmony is the result of the balance between **dominant** color (which has the greatest extension in the whole composition), **tonic** color (a vibrant color that gives hue to the whole) and **intermediate**

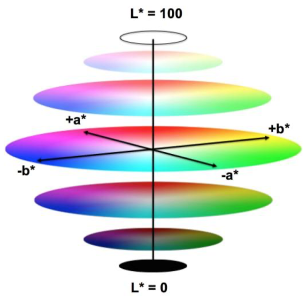


Figure 4.7: The LAB Color Model. The point at which the a* and b* axes cross, at the L* value of 50, is pure, balanced, neutral gray.

color (middle ground between dominant and tonic). All colors can be combined, but it does not mean that the result will be harmonic. When considering a possible color harmony between two elements, one should aim for the combination that provides greater visual conformity, arranging the colors as well as what the musicians do with musical notes - hence the terminological similarity between music and chromatic theory. When creating art, the chromatic harmony study of color is often considered as a limiter. Still, it is an ally that provides some knowledge to obtain different results [33]. There are some harmonic color schemes, all of which are based on the Color Wheel, as depicted in figure 4.8:

- **Monochromatic** - Colors composed of one hue and its brightness variation.
- **Complementary** - Colors that are opposite to each other on the color wheel.
- **Split Complementary** - This is a variation of the previous color scheme. In addition to the base color, it uses two adjacent colors to its complement.
- **Analogous** - Colors that are next to each other on the color wheel.
- **Triadic** - Colors that are evenly spaced around the color wheel.
- **Tetradic** - Four colors arranged into two complementary pairs.

4.3 Shape

When creating an image, a painting or a drawing, usually a two-dimensional space is generated by a shape in two dimensions: height and width. In the study of art, a shape is defined as *"an enclosed space, a bounded two-dimensional form that has both height and width"* [16]. As described in Section 4.1 shapes are one of the seven elements of art [15], the "building blocks" that artists use to create images. Their boundaries are defined by other elements of art, such as lines, colors, textures or value - with highlights and shadows, which can turn a shape into an illusion of its three-dimensional form.

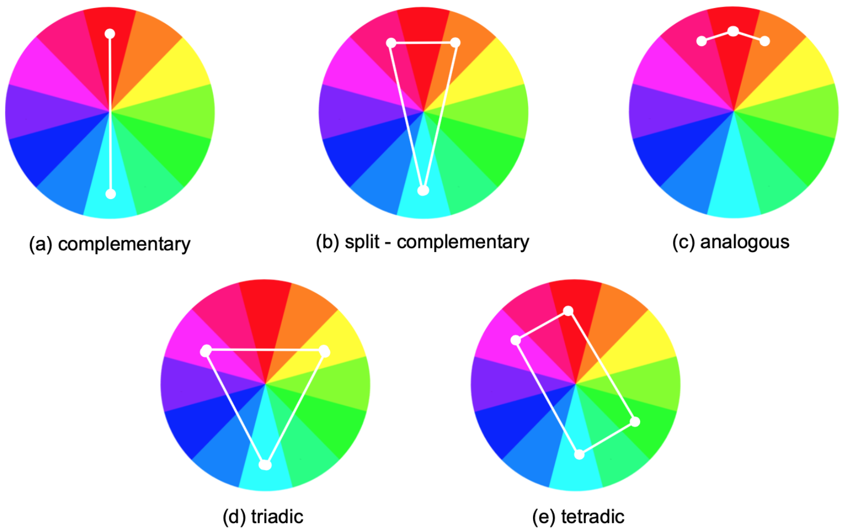


Figure 4.8: Color Schemes.

Line and shape are two elements in art that are nearly always used together. In the most basic approach, a shape is created when a line is enclosed - while the line forms the boundary, the shape is the form circumscribed by that boundary. For example, three lines create a triangle, while four lines can make a square. However, shapes are often divided into two categories:

- **Geometric** - those that are defined in mathematics and have common names - circle, triangle, square, polygon, and so forth. These shapes have clear edges or boundaries, and when physically drawing, artists often use tools as protractors and compasses to create and make them mathematically precise. The canvas is typically rectangular, implicitly defining the clear edges and boundaries of a painting or photograph [16].

- **Organic** - while geometric shapes are well-defined, organic shapes are just the opposite - free-form shape as what is usually found in Nature, when organic shapes can be as amorphous as a cloud or as precise as a leaf. These shapes are individual creations of the artists - they have no name, no defined angles, and no tools that support their distinct creation.

In the end, shapes are the element of art that is central to most artworks, as in the case of abstract art. They are everywhere, and all objects have it, whether they are Geometric or Organic [16].

4.4 Other Visual Features

According to the Seven Elements of Art [15], apart from color and shape, there are other five elements that can be used when creating an image artifact. These elements are:

- **Line** - defined by a point moving in space between two points whereby we can visualize stroke movement - may be two or three-dimensional, descriptive, implied, or abstract. It describes one outline, capable of producing texture.
- **Space** - it refers to the perspective (distance between and around) and proportion (size) between elements and how their relationship with the foreground or background are perceived.
- **Value** - refers to the degree of perceivable tone's lightness within an image. The value element is compatible with luminosity and refers to the lightness or darkness of tones of colors - the image's contrast. While white is the lightest value, black is the darkest, and the halfway value between these extremes is called the middle gray - see figure 4.2.
- **Form** - a three-dimensional element of art that encloses volume - includes height, width and depth (as in a cube, a sphere, a pyramid, or a cylinder). The form is often used when referring to physical works of art, such as sculptures.
- **Texture** - describes the surface quality of the artifact, related to the type of lines used.

These elements are considered essential to consider when creating art. First, it is impossible to create art without using at least a few of them [15]. Second, by knowing all these elements, people become able to describe what an artist has done, analyze what is going on in a particular piece, and communicate thoughts and findings using a common language known by everyone [15].

5

Approach and Implementation

Contents

5.1 Approach

We aim at a system that is capable of generating images inspired in music. To create an analogy from one domain to the other requires defining our starting language. However, before doing so, we needed to start learning music and its theory to understand how it is generally structured, in such a way we could analyze it and extract all the possible information regarding its composition. Having all the needed knowledge, mostly described in Section 3, we needed to restrict our approach regarding our starting point, the musical domain, since it is such a vast field with so many different styles and types. In Western music, if we think of a symphony, for example, it is divided into movements. According to our approach, these movements can be divided into sections and continue the division until measures, or to individual notes (minimum unit). Since we aim at generating abstract images that represent the music as a whole, any simplification, abbreviation, summary (reduction, in musical language), or other simplification of the genre, not being written by the composer, does not seem to be advisable. This way, regarding the duration of the artifacts, to restrict the musical domain, we came across two possible solutions:

1. Reduce the music by measures, for example: take into consideration the first 32 measures of a music piece, since it is possible to perceive some structure in them, without getting the feeling that a cut in half occurred.
2. Take into consideration music that lasts between a specific duration, for example, music that lasts between three and five minutes at most.

Since we want to be able to perceive all the musical elements in the generated image without the perception of elements' saturation, we decided that our approach would take into account the second alternative. As such, we decided to analyze music pieces that last between three and five minutes at most, so that it is possible to represent the music's harmony, melody and rhythm in such a way that one can identify all the translated elements in the computer screen. Moreover, cutting the music pieces in measures implies attentive listening and specific knowledge and practice in music that we do not have.

We also decided to analyze MIDI files to access the music's essential elements, how they are presented, structured, and organized. These files already have the sheet music in a computationally processed way where it is easy to manipulate and retrieve its characteristics.

The first step towards creating an association between the two domains was to map the music's features into visual elements and concepts. From the start, we did not want a too precise approach based on **one-to-one** associations[1], in which we would be generating a representation of the sheet music. We aimed at obtaining the musical component through the analysis of music's features in a

[1] One-to-one associations generate visualizations of sheet music. For example, each musical note (minimal unit) would be converted into a visual element.

general manner, by considering either the music or each the music's measures as a whole. Any one-to-one association made throughout the cross-domain module would be overtaken in the GA module - Section 5.2.5. With this in mind, let us start with a high-level association between the three parts of a musical piece, melody, harmony and rhythm, and the two groups of image's elements, background and foreground.

As described by Braga, F. and Pinto, H. [5], if we listen to a melody without any harmony (chords), we may perceive a particular sensation or emotion. Once we listen to it with harmony, this sensation or emotion may vary drastically from the initial one. For example, the same melody may be perceived as happier if we play it with a major harmony, whereas with a minor harmony, it may be perceived as sadder. Considering the visual domain, if we think about the image's foreground elements, we will obtain different sensations associated with its characteristics, such as darker or lighter. However, once we add a context to these foreground elements, we get, once again, a much better-defined sensation. We can say that while **harmony** gives context to the **melody**, the **background** gives context to the **foreground**. As such, our approach's foundations lie in the association of the melody with the foreground, and the harmony with the background. There is still one part of the music piece left to map, the **rhythm**, that is related to the music's time signature and the *tempo*. Concerning both, we decided to associate them with the size of the image's elements, as well as with the overall luminosity of the background. These associations shall be later on explained in this same Chapter.

Having our high-level association between harmony, melody and rhythm, and all the image's features, we then need to obtain each of the image's elements from the music's structure. The foreground seemed the right place to start.

As previously mentioned, the image's foreground is the closest view plan for the viewer, composed of elements that have, according to our approach, a well-defined shape, color, size and position. Accordingly, we tried to associate these characteristics with the music's melody. The most basic analogy that can be created between the musical and visual domains is that a note can be translated into a color. This way, in our first attempt, it only seemed natural to associate each note's pitch frequency with the color's frequency in the light spectrum, with a magnitude order fourteen times lower, so that it was possible to compare both quantities on the same scale. The same note-to-color association was tried with the note's pitch frequency and the color's magnitude in the visual spectrum. However, soon we realized that these two solutions were not the best choice since there would be some pitch frequencies without a direct association to color's frequency or magnitude. Still, comparing frequencies from different domains, with a different order of magnitude or unity of measure seemed not to be the best choice. Many different types of translation could be created, even random ones. However, we decided to follow the most natural way of association between the musical pitches and the color of the elements: we overlapped the chromatic scale, described in Section 3, and the color wheel, described in Section 4, where each pitch from the

Figure 5.1: Overlapping the Chromatic Circle and the RGB Color Wheel.

chromatic scale is associated to one, and only one hue from the color wheel - both the chromatic scale and the color wheel have twelve fractions organized sequentially, as depicted in figures 3.3 and 4.3. The A pitch-class is commonly considered as a reference standard with a frequency of 440Hz, used to calibrate musical instruments: we defined it as the first pitch-class of the chromatic scale. Red can be considered the first hue of the color wheel, since it is the first color in the visible spectrum, and, in the HSL model, it corresponds to *hue* = 0. This way, translating one to the other seemed to be a good starting point - the overlapped circles and our chosen translation are represented in figure 5.1.

When we decided the association between the note's pitch and element's color hues, we realized that two other visual properties could be added to the hue to improve the color's quality definition: saturation and lightness - creating a tuple of three visual properties for each color - (*hue*, *saturation*, *lightness*). These three characteristics led us to have the HSL color model in our mind, in which for the same hue, and the highest saturation (100%), the lightness channel will give us how dark or how pure that hue is, ranging from black (lightness close to 0%) to the pure hue (lightness close to 100%). To achieve this goal, we needed then to associate another two musical features related to its minimal unit - notes - to the saturation and lightness of the hue taken from the color wheel. From the note's octave, that represents how high or low a note is, we decided that this could give us how light or dark one color is. A lower octave corresponds to a darker color, with low value, whilst a higher octave corresponds to a lighter color, with a higher value. Finally, the intensity of the color is determined by the volume of the note. Just like the volume represents the strength of a note, the intensity represents the strength of a color.

Having established a possible analogy between note and color, we proceeded to test it out. At that time, we did not have a way to determine the element's shape, nor their size or position. As such, the image was generated with a number of elements correspondent to the number of minimal music units (notes), each with a random shape, size and position. Every note was read sequentially from the input MIDI file and translated into a visual element, generating one image with a limited dimension of

1000×1000 pixels. This approach resulted in images that were too saturated in elements - therefore impossible to perceive any pattern or structure - since we had pieces of music with several quantities of elements for the defined screen size. With this, we reinforced that we would not be able to analyze music pieces that last for too long. Besides, we defined the default value of 1600×2700 pixels for the generated images[2], although several dimension sizes for the generated images were tested throughout different phases of the development. Besides, images rendered with low width and height values (10×10 pixels, for instance) would be too small to represent all its elements, while images rendered with high values would be too big for the size of one computer screen. Nonetheless, these dimensions can be defined by the user in the beginning of the system's execution, and there are no limitations regarding the screen where the images are displayed - further described in Section 5.2. These experiments also demonstrate the decision of restricting the length of the music pieces in analysis to music that lasts between three and five minutes is favorable. Besides, this choice would facilitate our evaluation phase.

To continue the association process, we decided to follow the typical musical organization of the elements, and analyze each input file as a set of parts, each with a set of measures, as explained in Section 3. Each measure is analyzed individually, and a set of elements is generated from it. One measure is composed of a set of elements that can be notes, chords, time signatures or *tempo*. At this point, we could only generate a sequence of elements taken from the melody of the music, with random position, shape and size - this was the beginning of the foreground of our images. However, the artifact was still very simple and therefore required some other features to be added, as the element's position, shape and size. Besides, we also needed to define one criterion for the level of transparency of each element - alpha value - since we wanted to perceive a clear visual distinction between overlapped elements, either in the foreground, and background.

At this point, we needed an association to define the **shape** of each visual element. At first, we tried to arrange a criterion related to the harmonics series[3] of each instrument found on the musical sheet. The task would be to analyze recordings of some instruments (MIDI renders, for example) and each harmonic's energy to dynamically build a dedicated shape to that instrument. One possible approach would be to assign each harmonic to an equilateral polygon with so many sides as to the multiple associated with that harmonic, thus generating a hybrid shape between the various polygons. Despite that this approach would allow generating different shapes from a list of numbers with an explicit criterion for the shape of the instrument, it involves signal processing methods that we did not want to follow due to time constraints.

Later on, we decided to search for an association that better approximates each instruments' family to a shape, one visual representation based on the sound to be heard, and what it mentally reminds us. We

[2]The decision of the dimensions was based on the size of the computer screen in which this system was developed on.
[3]Harmonic Series are the sequence of frequencies, musical tones, or pure tones, in which each frequency is an integer multiple of a fundamental - the lowest frequency of a periodic waveform.

developed an online platform to study and customize all the possible associations between instruments and shapes to help in this process - it is available on https://creativity-2020.nw.r.appspot.com. This platform allows to choose a piece of music to analyze from a list of diverse Classical music that compose our dataset - further explained in Chapter 6. Therefore, it is possible to listen to the music and customize all the associations between each musical instrument found on the sheet music and visual shapes. In the end, the platform generates the output image.

Several different associations were tested through the platform in one interactive process. One possible association between each instrument's family and shapes, tested through the platform, is represented in table 5.1, where each shape was defined using two elements of the Seven Elements of Art [15] - shape and texture - thus increasing the abstraction level. With this approach, we associated the Chromatic Percussion and Piano's family to a circle, since these families are composed of instruments that can produce any sound. However, the Chromatic Percussion family is associated with smaller circles because, according to our perception, the sound is "drier" than the Piano's family - we thought about the sound of the xylophone versus the piano. Organ's family reminded us of a rectangular shape since it has religious roots, that may be considered cold and rigid, just like the walls of an old church. We consider the Guitar, Bass and String's sound as fluid, reason why visually we thought of these instruments almost as stains of watercolors that mix all together: Rectangular and Circular Spots. Ensemble's family has a set of different instruments, so we needed to analyze it case by case. Its shape would be a combination of the instruments that make the sound up: we considered it as a ten plus sided polygon. Brass's family, in our perception, has instruments related to the royalty, as trumpets, trombones and tubas. We related them to irregular squares. Reeds, as oboes and bagpipes, are linked to nature and hunting in their genesis: we mapped them to regular squares. Pipes are mostly flutes, that reminded us of butterflies or birds, the reason why we associated them with perfect rhombus. All the remaining instruments, as in the case of a synthesizer, can sound to everything. We decided to associate them to triangles.

It is important to note that we made this association based only in our perception, and that a mapping between instruments and shapes that has a perfect rational is hard to achieve. Moreover, since this is a subjective task, there would always be instruments that do not fit in these rigid shapes based only on our perception.

In the end, regarding the shape's association, we decided to have **two** different versions for the output of our system:

- **Random Image**, where a random shape and texture were assigned to each instrument found in the sheet music.
- **Associated Image**, where the previous associations between instruments and shapes were assigned.

The random version was defined because we consider that it improves the diversity of the element's

Family Name	Shape
Piano	Circle
Chromatic Percussion	Smaller Circle
Organ	Rectangle
Guitar	Rectangular Spot
Bass	Bigger Rectangular Spot
Strings	Circular Spot
Ensemble	+10 Polygon
Brass	Irregular Square
Reed	Square
Pipe	Rhombus
Other	Triangle

Table 5.1: One possible shape association for each instrument family.

shape, and generates images that, at our discretion, are more aesthetically pleasing. Moreover, if we always assign each instrument to the same shape, there would be generated images that consist of elements that always have the same shape, with low variety and diversity, as in the case if we analyze pieces composed only by violins (or any other instrument in repetition).

Regarding the rhythm, time signatures, plus information about *tempo* (BPM) are associated with the note's duration to define the element's **size**. All these three concepts were combined in a mathematical formula to achieve the best size for all the elements that could fit in a screen without the perception of visual saturation:

- **Time signature** indicates how many beats are in each measure of music. As explained in Section 3.5, a piece with a time signature of $\frac{4}{4}$ has four quarter-note beats, while a $\frac{3}{4}$ meter has three-quarter notes, for example. This way, we use the inverse of the time signature fraction in the size's calculation: the higher the time signature in **fraction**, the smaller the size. With this, we intended to represent, per measure, all the elements in a responsive size: measures with lower beat's duration represent bigger elements, whilst measures with more beats are visually smaller.

- ***Tempo*** is the pace or speed of the music: higher *tempo* means a faster song, whereas a slower *tempo* means a slower song. In our approach, we use an inverse relationship: the faster the music (higher BPM), the smaller the associated elements.

- **Note Duration** represents how long one element lasts in time: longer notes represent bigger elements, whilst slower notes represent small elements.

These three concepts significantly influence the way we perceive a music piece, thus they will affect the way our image's elements are represented.

Finally, to define the **position** of each element in the foreground, we use information regarding the offset - element's relative location in the music - of each melody's event, either notes and chords. This way, to build a dedicated output in which the elements are sequentially organized, the generated image is vertically divided into as many sections as the number of musical measures. Then, each vertical section is horizontally divided into as many sections as the number of events in the respective measure. At this point, we realized that different elements could have the same x and y positions, as in the case of one chord analysis, in which three or more notes are sounded simultaneously. However, this is not an issue, since each element has different characteristics. This way, information regarding the music elements' offset was used to directly define the elements' position in the screen.

Having analyzed the music's melody that was associated with the foreground, we can now move on to the background. It was decided to relate the image's background with the music's harmony. This way, we wanted to analyze the harmony and then associate the respective background segment's characteristics. As previously mentioned, the background of an artistic image is the part of a painting representing what lies behind objects in the foreground, composed of elements that have, according to our approach, a well-defined shape, color and position.

According to what was explained in Section 3, one can analyze the music as a whole to obtain the most probable tonality or scale used, thus obtaining a sense of its harmony. Right from the first iteration, we decided that we wanted a full analysis related to tonality, to most maximize the possible associations between both domains. At first, we started to analyze measure by measure, thus get its chord's root note. However, since we did not use the tonality in which the measure is inserted on, we decided to discard this possibility and move on to another solution. This way, we continued the analysis of each measure to obtain its most likely tonality key, thus performing a more in-depth analysis measure by measure, even if the certainty factor related to the tonality analysis is under a certain threshold. Having this in our mind, for each musical measure, we obtained the most probable tonality key to associate with characteristics of the background.

As previously stated, the most basic analogy that can be created between the musical and visual domains is that a note can be translated into a color. This time, having a set of tonality keys that belong to the music's harmony, we decided to overlap the Circle of Fifths, described in Section 3, and the Color Wheel, described in Section 4, where each key from the circle is associated to one, and only one hue from the color wheel - similarly to what happens with the chromatic circle, both the circle of fifths and the color wheel have twelve fractions organized in a sequential way. Furthermore, as well as with the chromatic circle, we decided the A pitch-class with a major key as the starting point for the overlap with the first color of the visible spectrum: red. The overlapped circles and our chosen translation are represented in figure 5.2.

At that time, we had an association between each measure's tonality key and the respective hue from the color wheel. To improve the color's quality definition, we decided to add, once again, to other properties to the hue - saturation and lightness. Since each measure's harmony is composed of chords,

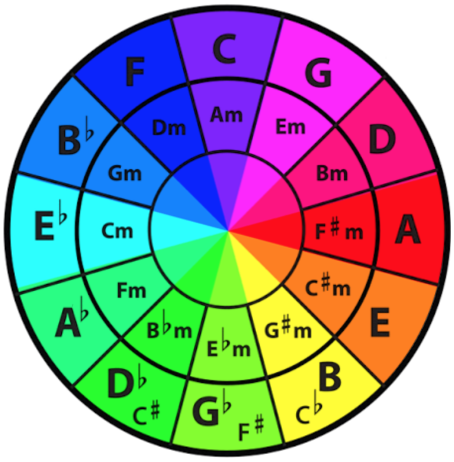


Figure 5.2: Overlapping the Circle of Fifths and the RGB Color Wheel.

we decided to get both the volume and each tonic note, also returning its octave. As before, the volume is associated with the color's saturation, while the octave defines the color's lightness and alpha value, thus establishing a possible association between chords and colors.

Having the image's foreground and background from the music's melody, harmony and the respective rhythm, we finally obtained the first two versions - **Random** and **Associated** Images. Besides these, we decided to implement a Genetic Algorithm to provide a way to search for better results while maintaining certain randomness that could lead to interesting results. Its implementation is described in the Section 5.2.5.

At this point, for each measure, we have its tonality key and the chords that make it up associated to one color characterized by three properties - hue, saturation and lightness. Having this, the remaining issues left to decide were the shape and location of each element that represents the measure's harmony. Our approach for the background considers each measure as a **vertical stripe** that is placed sequentially from left to right in the generated image, following the sequential way of writing a musical sheet in Western Countries. This way, having the measure's respective color, our background consists of equally sized vertical stripes, whose position is directly related to the measure's offset. For example, the first measure, with the lowest offset, is represented in the left side of the image, while the last measure, with the highest offset, is represented in the right side of the image. Each measure has its *tempo*, that directly influences how overlapped two stripes are, that is, the vertical stripe's **irregularity**: the higher the BPM, the less overlapped two stripes are, as well as the other way around. This decision was based on our visual perception of music: slower music (lower BPM value) are less "sharp" than faster music. Finally, to better differentiate the background from the foreground, we applied an **overall luminosity** filter, related to the *tempo* of the music: the lower the BPM, the darker the image, as well as the opposite, since slower music are more melancholic, therefore related to darker colors. In contrast, faster music have more energy, thus are related to lighter colors.

One interesting point to consider is one significant difference between the musical and the visual domains: whilst the latter does not have a specific order, a start and a finish, the former does. Whatever that order is, it influences the final results, especially the position of each element in the image, either foreground or background.

Having the image's foreground and background from the music's melody, harmony and the respective rhythm, we obtained the first two versions - Random and Associated Images - one with random associations between instruments and shapes, and other with predefined associations. At this point, we consider that we have a plausible association between the musical and the visual domains, represented in table 5.2. Besides this, we decided to use a Genetic Algorithm to provide a way to search for better results while maintaining certain randomness that leads to interesting results. The main issue when considering to use this Evolution technique - Section 2.2.1 - is the fitness function used to assess the generated individuals. In the first place we decided to evaluate the generated images according to its visual quality[4], using Image Quality Assessment (IQA) algorithms available online, that take an arbitrary image as input and output its **quality** score. However, since quality is a subjective matter, we needed to train - with human inputs - the algorithm about it, as well as about what good and bad quality are. Due to the implementation complexity, as well as lack of time, we decided to discard this approach and move on to our final one: image assessment according to color and music theory (Section 3 and 4), evaluating each visual element with its five nearest neighbours, accordingly - further described in Section 5. Besides, this approach is more related to our goals since we can assess both music and image characteristics, and their relations.

Finally, regarding the dimensions of the whole image, as previously mentioned, the 1600$\times$2700 pixels for the x and y values are not strict, since the user can manipulate them at the beginning of the run of our system. Besides, there is no limitation regarding the screens where the images are displayed.

[4]In the context of our work, we consider quality as a characteristic related to how distorted by blur, noise, watermarking, color transformations and geometric transformations one image is.

<table>
<tr><td colspan="3" rowspan="3"></td><th colspan="10">Image</th></tr>
<tr><th colspan="5">Background</th><th colspan="4">Foreground Elements</th><th rowspan="2">Dimension</th></tr>
<tr><th>Color</th><th>Shape</th><th>Position</th><th>Irregularity</th><th>Overall Luminosity</th><th>Color</th><th>Size</th><th>Position</th><th>Shape</th></tr>
<tr><th rowspan="7">Music</th><th colspan="2">Instrument</th><td></td><td rowspan="7">Vertical Stripes</td><td></td><td></td><td></td><td></td><td></td><td></td><td>X</td><td rowspan="7">1600 x 1700</td></tr>
<tr><th rowspan="2">Rhythm</th><td>Time Signature</td><td></td><td></td><td></td><td></td><td></td><td>X</td><td></td><td></td></tr>
<tr><td>BPM</td><td></td><td></td><td>X</td><td>X</td><td></td><td>X</td><td></td><td></td></tr>
<tr><th rowspan="2">Melody</th><td>Notes</td><td></td><td></td><td></td><td></td><td>X</td><td></td><td>X</td><td></td></tr>
<tr><td>Measure</td><td></td><td></td><td></td><td></td><td></td><td></td><td>X</td><td></td></tr>
<tr><th rowspan="2">Harmony</th><td>Chords</td><td>X</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td>Measure</td><td></td><td>X</td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
</table>

Table 5.2: Cross-Domain Associations.

5.2 Implementation

The first and significant aspect to consider when talking about implementation in a project like this is the programming language used. From the available options, the language Python[5] was chosen. This decision was based on the author's previous experience with it, mostly because it is a very versatile language with a great variety of manipulation tools. It has a complete standard library available, with the possibility to easily import other modules. Besides, it allows a multi-paradigm approach (Procedural, Functional and Object-Oriented).

The version used was the last one available at the start of the project, i.e., Python 3.8. Although the Python standard library is very extensive, other libraries were also used, since our work involves some specific domains:

- **Musical Domain**: to perform the extraction of all musical elements, so that these can be processed, the Music21[6] library was used. It provides a set of tools to answer questions from musicology using computers. Besides, it is used to study to compose music (both direct and algorithmically). It has operations to analyze each musical element, individually or collectively, and it is simple to use with an extensive documentation.

- **Visual Domain**: for the generation of all pixels in an image, the Pycairo[7] module was used. It is a library for drawing vector graphics that do not lose clarity when resized or transformed. It provides several tools and documentation, thus easy use.

Considering all the processes between the music and the image, we used other packages for some specific purposes. For handling arrays and some other collection of mathematical functions, we used Numpy[8]. It is a generally used package for scientific computing that enables numerical computing with Python, with an extensive documentation. For all the plotting purposes, we used the Matplotlib[9] package since it allows to create static, animated and interactive visualizations in Python. To define conversions between color values expressed in the RGB color space and other coordinate systems used in computer screens, we used Colorsys[10]. Finally, we used the Scikit-learn package[11] to implement neighbors-based learning methods.

Having referred the main packages used in our choice of programming language, we can now move into the following section, in which the architecture of the system is explained in detail.

[5] https://www.python.org
[6] http://web.mit.edu/music21/
[7] https://www.cairographics.org/pycairo/
[8] https://numpy.org
[9] https://matplotlib.org
[10] https://docs.python.org/2/library/colorsys.html
[11] https://scikit-learn.org

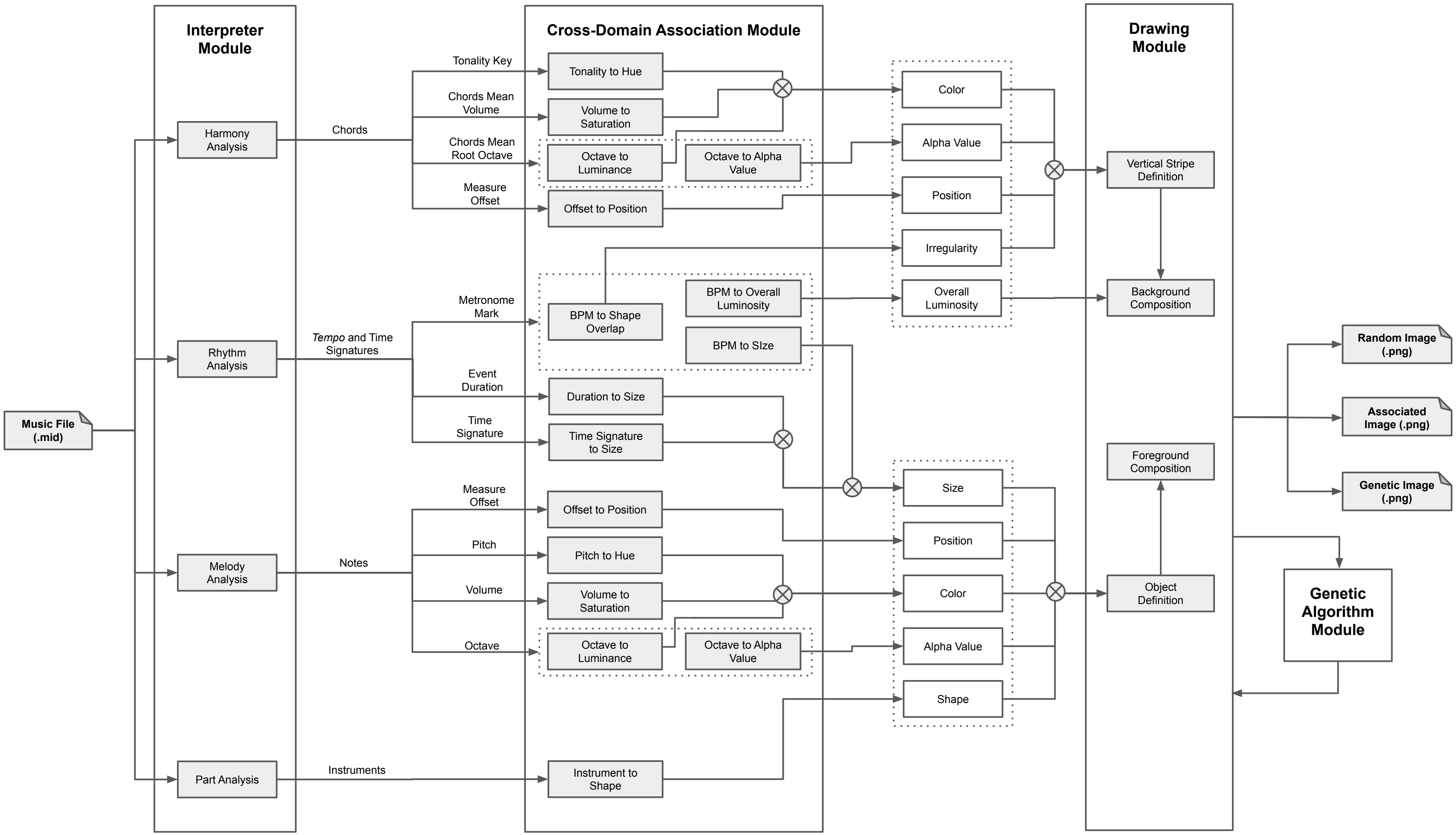


Figure 5.3: System Architecture.

5.2.1 System Architecture

Our system's architecture is divided into different modules to solve various problems individually, each containing different methods for a particular or specific goal. Figure 5.3 shows our general system's architecture, whose behavior and interactions are explained individually in the following sections.

In the beginning, our system receives a piece of music - **Interpreter Module**. In this module, the music (MIDI file) is processed and analyzed to obtain all its features. The results of this module are the music's harmony (chords), melody (notes) and rhythm (time signatures and *tempo*) features, as well as the timbre (instruments) that compose the score. Having these, we pass them on to the **Cross-Domain Association Module**, responsible for the mapping of the music's features to elements of the visual domain, except for the image's dimension and background's shapes, as explained in Section 5.1. The result of this module is a set of properties or values to be used in the generation - **Drawing Module**. At this stage, with all the necessary information, images are generated. Both the background and foreground were generated using music's harmony, melody and rhythm - while the former is composed considering the harmony, the later is composed considering the melody. Both consider the music's rhythm. Following this module, we have two possible scenarios: one is simply the output of the Random and Associated Images, while other is the **Genetic Algorithm Module**, responsible for the generation of the Genetic Image.

5.2.2 Interpreter Module

This module is responsible for extracting and processing the needed information from the music given as input. After receiving the MIDI file, we use the Music21 library to process it: a top-down analysis is done for each file, from its parts to its minimal unity, notes. Our system is organized in such a way that we first analyze music's Parts, then Measures and finally, measure's Events, in a hierarchical manner:

- When analyzing a **Part**, we get the correspondent Instrument name, as well as its program change[12].
- When analyzing a **Measure**, we get its offset.
- When analyzing an **Event**, four different scenarios can occur:
  - If we find an instance of **MetronomeMark**, it means that we are dealing with music's *tempo* for that measure. As such, we get its value (BPM).
  - If we find an instance of **TimeSignature**, it means that we found the music's time signature for that measure. Thus, we get its value.
  - If we find an instance of **Note**, **Chord**, or Rest, first we get and store its offset. Then, if it is a Note, we get its correspondent Pitch, as well as its Volume and Octave. If it is a Chord, as

[12] Program change is a MIDI message that identifies the instrumental sound that the MIDI device uses when it plays a Note.

well as with Notes, we get its duration and perform an analysis element by element - note by note that composes that chord - retrieving their characteristics.

Having extracted all the information regarding the musical parts, measures and events, we have found enough information for the music's melody and rhythm. To store all these feature's values, we used RecordType[13], that similarly to named tuples, it is a method for making mutable tuple classes that can be callable by a name or identifier. Finally, we used a Music21 available method to process and analyze the music's harmony, retrieving the most likely tonality key of each measure. This method is a process to make chords out of non-chords: *it is a powerful tool for reducing a complex score with multiple parts to a succession of chords in one part that represents everything that is happening in the score*. This way, having all the parts unified in just one, we can proceed to the measure's offset retrieval and tonality analysis. As explained before, this one part that represents the harmony has measures that are composed of chords. For each, we calculate its Volume and Root Octave, performing an average for that measure at the end of the correspondent analysis. Finally, using a method for tonality retrieval, we found out, for each measure, its most likely tonality key.

5.2.3 Cross-Domain Association Module

Every operation that is a translation from the music to the visual domains is implemented in this module, carefully explained in Section 5.1:

- Starting with the part analysis, we get the shape definition. As explained in Section 5.1 and table 5.1, two different approaches were considered. Regarding the **Random** Image, for each instrument found in the music, we decided to randomly assign a shape among 17 available: small circle, circle, circle spot, triangle, rectangle, rectangular spot, big rectangular spot, square, irregular square, rhombus, pentagon, hexagon, heptagon, octagon, eneagon, decagon and 10+ sides polygon. Regarding the **Associated** Image, the associations presented in table 5.1 were applied.
- From the music's offset to the element's visual position: we normalized the measure's offset to the x value, and the event's offset to the y value, according to the dimensions of the image.
- From note's pitch to the color's hue definition, among 12 available. The respective association between the Chromatic Circle and the RGB Color Wheel is available in figure 5.1. The resulting RGB Color serves as the Hue for the HSL Color Model.
- From note's volume - normalized between 0 and 1 - to color's saturation.
- From note's octave to color's lightness and alpha value. After some experiments, we consider seven different octaves, from the zeroth to the sixth, and lightness values ranged between 0.1 and

[13] https://pypi.org/project/recordtype/

Tempo (BPMs)	Overall Luminosity	Stripe's Irregularity
$<$60	0.0	5.0
60-69	0.1	4.5
70-79	0.2	4.0
80-89	0.3	3.5
90-99	0.4	3.0
100-109	0.5	2.5
110-119	0.6	2.0
120-129	0.7	1.5
130-139	0.8	1.0
140-149	0.9	0.5
$>$150	1.0	0.0

Table 5.3: Association from *tempo* to overall luminosity and stripe's irregularity.

0.7. While the former is a direct translation, the later applies an inverse operation: the higher the octave, the less solid (transparent) the color is: $alpha = 1 - lightness$.

- From note's duration, time signature and *tempo* to the element's visual size: for each element, considering what is explained in Section 5.1, we apply the following equation:

$$size = \left(Duration \times \frac{1}{Tempo}\right) + \left(\frac{1}{Tempo} \times \frac{1}{TimeSignature}\right) \tag{5.1}$$

- From music's *tempo* to background's overall luminosity and stripe's irregularity by overlap - table 5.3. The music's *tempo* range from 60 (slow) to 150 (fast), while the luminosity from 0 (black) to 1 (white), and irregularity from 5 (regular) to 0 (irregular).

- From tonality key to the color's hue definition, among 12 available. The respective association between the Circle of Fifths and the RGB Color Wheel is represented in figure 5.2. The resulting RGB Color serves as the Hue for the HSL Color Model.

- From chord's root note volume to color's saturation. For every chord found on each measure, we extract its root note to get its volume. Then, having all, we calculate an average for the volume of that chord.

- From chord's root note octave to color's lightness and alpha value. For each chord on each measure, we extract its root note to get its octave. Then, having all, we calculate an average for the predominant octave of that chord.

The rationale behind these associations is carefully explained in Section 5.1.

5.2.4 Drawing Module

In this module, the generation of the abstract image is performed. The task is divided into two parts: the **foreground** composition and the **background** composition.

First, considering the foreground composition, five features need to be generated for each element: shape, size, color, position and alpha value. For the shape, as explained, we use the two shape distributions from the previous module, generating 17 different shapes defined by mathematical equations applied with texture (deformation) filters:

- To define a circle, we specify its center point $(x_c,\ y_c)$ and, using a loop for each circumference degree (out of 360), i, we define its describing points according to equation 5.2. By default, n is 360, and represents the visible angle of the circumference.

$$circle(x_c, y_c, n) = \begin{cases} x = x_c + r \times cos(\frac{i \times \pi}{180}) \\ \\ y = y_c + r \times sin(\frac{i \times \pi}{180}) \end{cases} \tag{5.2}$$

- To define a regular polygon, we specify its starting point, $(x_0,\ y_0)$, and the number of sides, s. Then, using a loop for each side, i, we define its describing points according to equation 5.3.

$$polygon(x_0, y_0, s) = \begin{cases} x = x_0 + sin(\frac{2\pi}{s} \times i) \\ \\ y = y_0 + cos(\frac{2\pi}{s} \times i) \end{cases} \tag{5.3}$$

- To define irregular shapes, we define either a circle or a polygon, and apply a deformation to most maximize the texture filter - for example, by iteratively finding and adding the middle point of each polygon's side, and applying an alpha value, one rigid shape may look like a stain.

All the remaining associations for the foreground, both color, position, and size were already translated and explained from the previous module. All the features regarding each visual element were stored in a Numpy array using Record Types. This way, having described all the needed features to define an element, we can move on to the background composition.

At this point, considering the background's approach explained in Section 5.1, having the color, position, irregularity and overall luminosity features directly extracted from the previous module, we have all the necessary features to iteratively draw the final image.

Using Pycairo, we start by creating a context[14] and an image surface with the desirable size of the image - 1600×2700 pixels as default. Then, we continue by defining the background as a black rectangle with the overall luminosity filter applied. Finally, using a linear gradient, we draw the vertical stripes considering the respective irregularity, color and position features. To fill out the foreground,

[14]Context is the main object used when drawing with Pycairo (to draw an image, we have to create a Context, set the target surface, and drawing options for the Context).

since each element has already all the necessary features attached, the system sequentially traverses the Numpy array and, considering each element, draws it using Pycairo drawing methods. Finally, we save the final Random and Associated Images to a separate $.png$ file, the first two **system's outputs**.

5.2.5 Genetic Algorithm Module

As explained in Section 2.2.1, a Genetic Algorithm (GA) is a search technique analogous to the process of natural selection and sexual reproduction. Since we want to improve the quality and novelty of the generated images, this algorithm was mainly applied to the visual domain, besides we consider some concepts of music harmony in the generated elements.

There are different types of representations for genes - binary, decimal, integer -, each treated differently. As well as with representation, there are different types of **mutation** - bit flip, swap, inverse, uniform, Gaussian, shrink, and others - and **crossover** (blend, one point, uniform, and others). In our implementation, we use a decimal representation for genes, fitness proportionate selection, one-point crossover, and uniform mutation [29]. The following sections explain in detail how we implemented the GA: the initialization, selection, crossover and mutation operations, as well as the fitness functions we considered to evaluate each individual - figure 2.1.

5.2.5.A Initialization

Initially, an arbitrary number of possible solutions for the problem - individuals - were generated, creating the initial population that shall hold within a wide range of solutions to explore all the different possibilities over the run of the algorithm. In our case, an individual is an image artifact which contains a set of elements in the background and foreground, each with its own characteristics.

The number of possible individuals per population depends on the dimensions, scaling factor, crossover and mutation rate, as well as the nature of the problems to be solved. Since we want to improve the quality of the abstract images by maximizing the variety of generated artifacts, we tried several population sizes. After experiencing large population sizes, above 100 individuals, the algorithm took a very long time to run. With small population sizes, under 20 individuals, the algorithm converged in fewer generations, which seemed to be a good issue to our implementation. However, according to [36], good population size is about 20-30 individuals, even if depending on the complexity of the gene, sometimes sizes 50-100 are reported as best. This way, we decided to define the population size as 24 individuals, where the first two are the Random and Associated Images, previously generated. The other 22 individuals were created from the previous images, where each foreground's element was shuffled according to its x and y location, creating our first population of 24 individuals. Having generated the first population, the following were created through genetic operators explained in the following sections.

5.2.5.B Selection

Selection implements the survival of the fittest individuals according to predefined fitness measures - in our case, mainly according to **color** and **music** theory. To select each pair of individuals for mating, and propagating their features to the next generation, Roulette Wheel Selection was chosen. With this technique, every individual can become a parent with a probability proportional to its fitness. Therefore, higher-fitness individuals have a better chance to survive and reproduce, while lower-fitness ones are more likely to disappear. To guarantee some level of quality on the later iterations, we use an Elitism factor - for each new generation, we pass exactly four individuals from the previous one. Therefore, if new individuals do not have better fitness values than their parents, then at least we maintain the highest values of the prior generation [29].

5.2.5.C Crossover

When creating a GA, crossover is an operation performed to create an individual that is constituted by a part of one individual and another part of a second individual. This operation is done by selecting a part in each of the two original individuals, and switch those parts between them, creating then two new individuals. According to [36], crossover rate generally should be high, about **80%-95%**, and since we were using **single-point crossover**, the crossover takes place in the point that divides the solutions into two equal halves. In our implementation, on each pair of selected individuals, there is a **90% probability** of occurring crossover, which can be one of the following specific types:

- **Color Crossover** - where only the elements' colors of one individual are crossed with the elements' colors of another individual.
- **Position Crossover** - where only the elements' position of one individual is crossed with the elements' position of another individual.
- **Shape Crossover** - where only the elements' shape of one individual is crossed with the elements' shape of another individual.
- **Size Crossover** - where only the elements' size of one individual is crossed with the elements' size of another individual.

It is important to note that for each pair of selected individuals, only one of the previous crossover types occurs.

5.2.5.D Mutation

Mutation is when a part of an individual is changed by another random part, as well as with gene's mutations.

According to [36], mutation rate should be low, where the best rates reported are about **0.5%-1%**. However, in our case, we defined the mutation rate to be a bit higher since many experiments were done with lower values that did not allow the algorithm to converge. In our case, mutation involves changing one element characteristic, and similarly to crossover, several types may occur:

- **Position Mutation** - with a 10% chance to occur, the x and y locations are changed to random values within the dimensions of the image.
- **Offset Mutation** - similarly to the previous, this mutation changes the x or y values to positions either horizontally or vertically aligned with that one. This mutation has a 5% chance to occur.
- **Color Mutation** - with a 10% chance to occur, the Hue component of the HSL Color Model is changed to a random hue among the 12 available on the Color Wheel.
- **Shape Mutation** - with a 5% chance to occur, the element's shape is changed to another one, among the 17 available.

Having explained all the operations applied to generate new populations, we can move on to the explanation of the fitness assessment. This requires some further explanation regarding implementation, since we evaluate images according to their elements, considering music and color harmony.

5.2.5.E Fitness Function

At this stage, the new individuals need to have their fitness calculation, which is a sum of different measures:

- **Location on the canvas** - if the position of each element lies between a distance of 10 pixels from its original exact position, then:
  - If the x position is within the defined range, $+8.5$ is added.
  - If the y position is within the defined range, $+5$ is added.

  This assessment is based on the fact that the canvas' default width - 2700 pixels - is 1.7 times larger than its height - 1600 pixels.
- **Shape's Polygon** - polygons with a higher number of sides resemble to circles. Since we want to visually maximize the diversity of the generated elements, then for each element, if its shape is:
  - a small circle, then $+7.5$ is added.
  - a cicrle, then $+9$ is added.
  - an octagon (8 sides polygon), then -5 is added.

- a nonagon (9 sides polygon), then -6 is added.
- a decagon (10 sides polygon), then -7 is added.
- $+10$ sides polygon, then -10 is added.
- for all the remaining shapes, $+6$ is added

- **Shape's Association** - if the element's shape follow the previous association in table 5.1, then $+10$ is added.

At this stage, regarding what was explained in Section 5.1, to make comparisons between elements, we first started to build, for each element, their nearest neighbours tree. The choice of neighbors search algorithm[15] is controlled through the keyword 'algorithm', which can be the *Ball Tree* or the *KD Tree*, for example. In our case, we defined the algorithm parameter with the default value 'auto', as it attempts to determine the best approach from the training data. This way, for each element, we find the five nearest neighbours based on the 'euclidean' distance metric. Then, for each element and its nearest neighbours, we applied the following measures:

- **Delta E** - to calculate the perception humans have between two colors, we use the Delta E distance metric to directly compare colors based on the LAB Color Model - figure 4.7. This way, for each element and their neighbour, if the distance is:
  - above 15, then $+15$ is added.
  - under 15, then $+delta_e$ is added.

- **Color Harmony** - according to Section 4.8, there are some basic color chords based on the color wheel. Then, for each two nearest elements, we calculate their color harmony according to the following measures:
  - if their colors are complementary, $+10$ is added.
  - if their colors are split complementary, $+7.5$ is added.
  - if their colors are analogous, $+5$ is added.
  - if their colors are triadic, $+2.5$ is added.
  - if their colors are tetradic, $+2.5$ is added.
  - if their colors have no harmony between them, -1 is added.

- **Music Harmony** - having shuffled the position of each element in the canvas, we decided to assess two elements considering their musical harmony. This way, for each two nearest elements, we calculate if their pitches are next to each other on the Chromatic Circle - figure 4.2.2. Then, if two elements are in the same x and y location, $+5$ is added. Otherwise, $+8.5$ is added.

[15] https://scikit-learn.org/stable/modules/neighbors.html#nearest-neighbor-algorithms

On the first hand, we give score to nearest elements that have harmony between their colors. On the other hand, we remove score if their color's perception to the human eye are visually close. Therefore, we maximize the color's diversity of the generated elements considering its harmony and visual perception, which may lead to interesting results. In the end, we decided to consider the element's pitch harmony, since each element was shuffled at the beginning of the algorithm - if two elements visually closer to each other have harmony between their pitches, we give score according to their location.

Finally, having explained all the GA parameters, it is essential to note that we decided that our algorithm finishes after 300 generations have been calculated. The final solution is the best population with the fittest individual, that is returned to the Generator Module - Section 5.2.4, thus generating the Genetic Image.

6

Case Study

Contents

This project was divided into two main phases, the first, where the **Random** and **Associated** Images were generated, and the second, where the GA was added to create the **Genetic** Image. Although we could evaluate each generated image in two different stages, we decided to assess our system's behavior in its final version[1], including the GA. It is important to note that both the Random and Associated Images were generated in parallel, while the Genetic was generated considering these two, at a later stage. The time needed to run the algorithm depends on the size, duration, and complexity of the music file. For the first two versions, the running time is entirely related to the duration of the music itself, since the slowest part of the whole generation process is the iteration of all the elements in the music. The time taken to run the GA is related to the spatial position of each element in the canvas, since we are performing an in-depth search considering the closest neighbors of each visual element. Therefore, its complexity is $O(n^2)$ - where n is the number of elements that constitute the image - that may result in several hours of running time.

6.1 Dataset

We restricted our input domain to **Classical** music, using MIDI files from different periods and composers. Besides, this music style is a form of art produced in the traditions of Western culture, noted for its diversity and development of sophisticated forms of instrumental music - symphony, concerto and sonata - which may be interesting to our approach.

As explained in Section 5.1, regarding the musical domain, due to the considerable amount of elements to analyze, as well as the computational time to run the algorithms, we chose music files that have a duration between three and five minutes. Thus, all the necessary features can be extracted and visually represented in a screen. Most of the music files were downloaded from the *Kunst der Fuge*[2] website, which is the largest resource of Classical music in *.mid* files, with more than 19000 music from more than 1000 authors. Having all, we used *MuseScore*[3] software to visually analyze each musical score and get a sense of its structure.

Considering the Classical style, we analysed at least three samples from the *Baroque*, *Classic*, *Romantic* and *Modernism* periods, all from different composers. It is important to note that, our approach is independent of the musical style, and any could have been considered, depending on the online availability of the correspondent MIDI files.

All the fourteen MIDI that compose the dataset are represented with the respective name, composer, period and style in table 6.1. Although we have executed the system with each of the MIDI files that compose our dataset, in the following subsection we provide a detailed analysis on four of them, whose

[1] For the purpose of the evaluation process, we call the generated images - Random, Associated and Genetic - as our three versions for each musical input.

[2] http://kunstderfuge.com

[3] https://musescore.com

<table>
<tr><th>Style</th><th>Period</th><th>Composer</th><th>MIDI File</th></tr>
<tr><td rowspan="14">Classical</td><td rowspan="4">Barroque</td><td rowspan="2">Bach</td><td>Cantata BWV 150, 2nd movement</td></tr>
<tr><td>Cantata BWV 4 Christ lag in todesbanden</td></tr>
<tr><td rowspan="2">Handel</td><td>The Royal Fireworks Suite - 3. La Paix (HWV 351)</td></tr>
<tr><td>Concerto No. 1 - 2. Largo (HWV 312)</td></tr>
<tr><td rowspan="4">Classic</td><td rowspan="2">Mozart</td><td>Symphony No. 40 - 3. Menuetto and Trio; Allegretto</td></tr>
<tr><td>Symphony No. 41 - 3. Menuetto</td></tr>
<tr><td rowspan="2">Haydn</td><td>Symphony No. 94 - 4. Finale: Allegro molto</td></tr>
<tr><td>Symphony No. 94 - 3. Minuettoo e trio: Allegro molto</td></tr>
<tr><td rowspan="3">Romantic</td><td rowspan="2">Beethoven</td><td>Symphony No. 8 - 2. Allegretto scherzando</td></tr>
<tr><td>Symphony No. 2 - 3. Scherzo - Allegro</td></tr>
<tr><td>Schumann</td><td>Kinderszenen / Scenes from Childhood Op. 15</td></tr>
<tr><td rowspan="3">Modernism</td><td rowspan="2">Stravinsky</td><td>L'Oiseau de Feu / The Firebird 1. Introduction</td></tr>
<tr><td>The rite of spring, 2nd movement (1st part)</td></tr>
<tr><td>Debussy</td><td>Suite Bergamesque, 1st movement</td></tr>
</table>

Table 6.1: Input Files from different Classical Periods.

results were evaluated through online surveys. The evaluated music files and respective translated versions are:

1. Georg Handel - *Concerto No. 1 - 2. Largo*
2. Georg Handel - *The Royal Fireworks Suite - 3. La Paix*
3. Wolfgang Mozart - *Symphony No. 40 - 3. Menuetto and Trio; Allegretto*
4. Ígor Stravinsky - *The Firebird - 1. Introduction*

6.2 System's Process and Output

All source files can be found in the "musics" folder, inside the main directory of the project. This system was implemented and executed with the PyCharm Environment[4] in MacBook Pro Catalina. To run the system, with Python 3.8 interpreter, the run.py file should be executed, with the respective music name and desirable dimensions for the output images. After starting the execution of the process, the following steps are performed sequentially, considering what was explained in Chapter 5:

1. MIDI processing and features extraction.
2. Features' association with visual elements and their characteristics.
3. Generation of the Random and Associated Images (first and second output).
4. Execution of the Genetic Algorithm over the previous artifacts.

[4] https://www.jetbrains.com/pycharm/

5. Generation of the Genetic Image (third output).

When the process is finished, the three image versions can be found in the root directory of the project, whose names are directly related to the version to which it corresponds: “music_name-random.png”, “music_name-associated.png” and “music_name-genetic.png”.

6.3 Evaluation

Having all the outputs of our system, we moved on to the evaluation phase, where the most important aspects to consider are our objectives. Recalling the Ritchie Model of Assessing Creativity - presented in Section 2.1 - we can analyze our system to determine if it behaves creatively according to the novelty, quality and typicality of the generated images. However, this method of assessing creativity mainly focus on the system’s behaviour - on the path to get the output - and not on the output itself [42]. Since our goals are to verify if the generated images are abstract and related to the music, we decided to focus on the assessment of the system’s input and output. Besides, we consider that this assessment will indirectly lead to the evaluation of the system’s behaviour.

Over the development of our system, we noticed that images were visually represented in slightly different color levels, depending on the computer screen in usage. Considering that the representation of a color model depends on the computer’s graphic card, as well as on the screen’s luminosity, we decided to apply two filters to each generated image. With this, we hope to improve the quality of the generated artifacts, as well as to minimize its luminosity levels fluctuation on different screens. Therefore, for each output, we applied an image normalization operator that finds the exact minimum and maximum color values in the image, applying then a level operator to stretch the values to the full range of values. After some experiments, we realized that if a second filter related to the image’s contrast were applied, the results would be even more uniform among all screens. Thus, the second contrast filter[5] was applied together with the previous to enhance the intensity differences between the lighter and darker elements of the generated images, overcoming problems related to fluctuating color levels. It is important to note that these filters were applied with the same parameters to all the generated images, so that the images’ color levels are visually equal among all screens.

To carry out the evaluation process, we needed to decide which set of images we would like to assess. Therefore, to increase the output’s diversity, we decided to run the algorithm over four different music from distinct periods and composers. This way, the output images are visually more distinctive and heterogeneous, exploring all the potential of our system.

We elaborated four different surveys with four alternated music. Each one has two music and the

[5]The contrast filter, as well as the normalization operator, is fully described on https://imagemagick.org

respective set images to be analyzed[6]:

- **Survey 1**: Concerto Handel & Stravinsky - we evaluate two sets of images translated from music of the Baroque (Handel - *Concerto No. 1 - 2. Largo*) and Modernism (Stravinsky - *The Firebird - 1. Introduction*) periods.
- **Survey 2**: Concerto Handel & Royal Fireworks - we evaluate two sets of images translated from the Baroque period (Handel - *Concerto No. 1 - 2. Largo* and Handel - *The Royal Fireworks Suite - 3. La Paix*).
- **Survey 3**: Mozart & Stravinsky - we evaluate two sets of images translated from the Classic (Mozart - *Symphony No. 40 - 3. Menuetto and Trio; Allegretto*) and Modernism (Stravinsky - *The Firebird - 1. Introduction*) periods.
- **Survey 4**: Royal Fireworks & Mozart - we evaluate two sets of images translated from the Baroque (Handel - *The Royal Fireworks Suite - 3. La Paix*) and Classic (Mozart - *Symphony No. 40 - 3. Menuetto and Trio; Allegretto*) periods.

All the output images used in the evaluation process can be seen in appendix A. Having these, we equally assess four different sets of images distributed in the four surveys described above. The survey starts by asking the age group of people $< 18; 18-29; 30-49; 50-59; > 60$; and the gender - masculine, feminine, or other. Both these questions have the “Prefer not to answer” option, since there may exist people who do not feel comfortable in sharing their personal information. Finally, the survey asks how often do the participants go to exhibitions, online or before the COVID-19 pandemic - Never; Once a year; Once a month; Every week.

Afterward, for each set of generated images, the survey starts by showing each version and asks to describe them, using three own sentences (open answer). This allows the participants to express what they exactly see and feel, thus not restricting their answers. Then, based on the set of images, the survey asks if users consider them to be abstract images - yes or no - and to rank them from the most preferred, to the least preferred. With these two questions, we intend to evaluate the images itself, while with the description, we intend to get a more subjective opinion.

After that, in a different page, for the same set of images, the survey asks to describe each version according to predefined adjectives: Exciting; Smooth; Happy; Enjoyable; Surprising; Contemptible; Sad; Aggressive; Disgusting; Boring; Angerful; Fearful. This choice of adjectives were based on our perception from the generated images, as well as on the Ekman’s research on universal emotions [13] - Anger, Contempt, Disgust, Enjoyment, Fear, Sadness and Surprise - allowing then to perform an emotional categorization. The questions regarding the images’ description allow a wide range of answers, thus

[6] It is important to note that, depending on the survey, the sequence in which the images are shown changes.

a full analysis of the image's perception and respective emotions can be performed. Besides, we ask them on different pages to avoid the users to be biased from the previous adjectives.

The survey moves then to the part related to the musical domain - the respective music is shown without any reference to it being used as inspiration for the image generation. Then, the survey repeats the process regarding the image description and asks to describe the musical artifact using three own sentences. Afterward, in a new page, it asks to describe the music according to the previous predefined adjectives: Exciting; Smooth; Happy; Enjoyable; Surprising; Contemptible; Sad; Aggressive; Disgusting; Boring; Angerful; Fearful. With this, we intend to make the users to carefully listen to the music and obtain a general description of it. After that, the survey shows the most and least preferred images, and for each, ask for the following questions:

- "How would you rate this image?" 1 (Do not like it) - 5 (Really like it)
- "Do you think this image is related to the music?" 1 (Not at all) - 5 (Totally)

With this, we intended, once again, to evaluate the image itself, as well as to perceive it the audience recognize any inspiration from the music. Finally, we come to the part regarding the relation between image and music. Here, the survey cites the following sentence - "All the images were created having the previous music as inspiration." - and asks if the user agrees with it: 1 (Not at all) - 5 (Totally).

In the following section, we both present and discuss all the results that we got from each survey.

6.4 Results and Discussion

To assess each music-image pair, we equally distributed the four surveys to heterogeneous groups of people. In total, 93 people - 62 female and 31 male participants - were questioned through all the surveys. Starting by the age group, the main one, for all the surveys, was $18 - 29$ - all the results from each survey, in percentage, can be seen in figure 6.1.

Regarding how often participants go to exhibitions, online or before the COVID-19 pandemic, the majority answered "Once a year", which was what we were expecting. The obtained results, in percentage, for each survey, can be seen in figure 6.2. It is important to note that, from this moment on, our analysis concerns on the music-image itself, and not on the survey in which they were asked.

Having all the information from the participants, we move on to the image description. However, since we noticed that the descriptors used by the participants in each version of the generated image are, someway, very similar, we decided to focus our attention on the most and least preferred images. This way, we are going to present the statistics for each pair music-image, and then relate them to the adjectival description participants had on their perceptions. Besides, by doing so, the analysis becomes more concise and less repetitive.

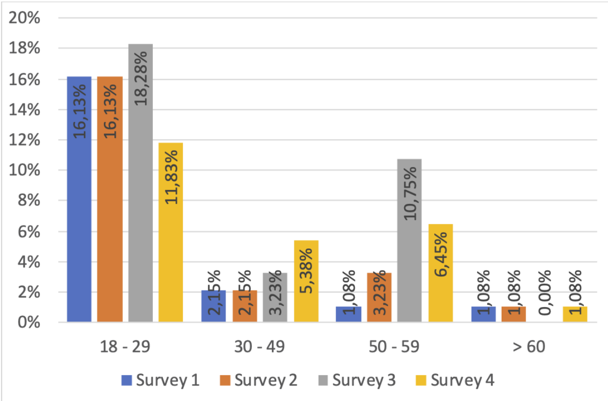


Figure 6.1: What is your age group?

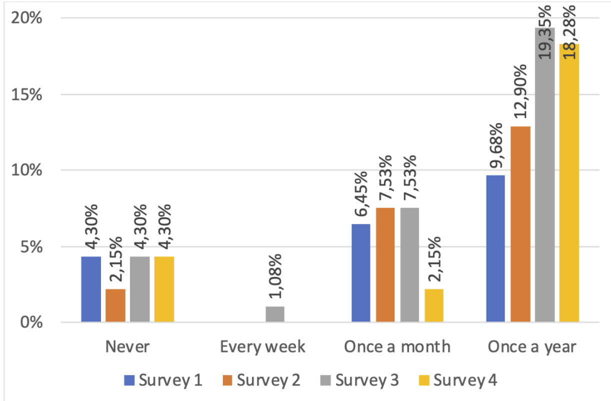


Figure 6.2: How often do you usually go to exhibitions?

Question	Music	Image version			Total
		Random	**Associated**	**Genetic**	
Which image version do you prefer?	Handel - *Concerto No. 1* *2. Largo*	**15.00%**	**7.50%**	**77.50%**	**100%**
	Handel - *The Royal Fireworks Suite* *3. La Paix*	**34.09%**	**38.64%**	**27.27%**	**100%**
	Mozart - *Symphony No. 40* *3. Menuetto and Trio; Allegretto*	**22.64%**	**28.30%**	**49.06%**	**100%**
	Stravinsky - *The Firebird* *1. Introduction*	**32.65%**	**42.86%**	**24.49%**	**100%**
	Total	**26.34%**	**30.11%**	**43.55%**	**100%**

Table 6.2: Which image version do you prefer?

All the participants ranked the three image versions, and the relative frequency for each music and respective preferred image version can be seen in table 6.2. From this, we conclude that the most preferred image version, in the majority, is the Genetic one. However, depending on the music, we analyze that this choice may vary between the Associated and Genetic versions. This goes in line with what we were expecting, since the Genetic image evolves from the Random and Associated images, in one iterative process that searches for better results.

The mean, mode, and standard deviation for the results of the most preferred image can be seen in table 6.3. It is important to note that the scale in both cases goes from 1 (lowest) to 5 (highest), which means that the closest to 5, the better the results are. Regarding the first question (How would you rate this image?), the results are very positive since the mean and median in all the generated images is above 3. Besides, the mode in three of the four sets of images is 4, and in the Stravinsky case, is 5, which means that for the preferred version, participants generally really like it.

For the second question (Do you think this image is related to the music?), the answers are not so constant. By looking at the results, we see that the statistics are, in some cases, under 3. Since the scale goes from 1 (no relation) to 5 (totally related), 3 indicates the middle, meaning that participants

Question	Music	Mean	Median	Mode	Standard Deviation
How would you rate this image?	Handel - *Concerto No.* 1 2. *Largo*	**3.67**	**4**	**4**	**0.76**
	Handel - *The Royal Fireworks Suite* 3. *La Paix*	**3.70**	**4**	**4**	**1.09**
	Mozart - *Symphony No. 40* *3. Menuetto and Trio; Allegretto*	**3.68**	**4**	**4**	**0.94**
	Stravinsky - *The Firebird* *1. Introduction*	**3.16**	**3**	**5**	**1.45**
Question	**Music**	**Mean**	**Median**	**Mode**	**Standard Deviation**
Do you think this image is related to the music?	Handel - *Concerto No. 1* *2. Largo*	**2.38**	**2**	**2**	**1.27**
	Handel - *The Royal Fireworks Suite* *3. La Paix*	**2.93**	**3**	**4**	**1.21**
	Mozart - *Symphony No. 40* *3. Menuetto and Trio; Allegretto*	**3.23**	**3**	**4**	**1.09**
	Stravinsky - *The Firebird* *1. Introduction*	**3.33**	**3**	**2**	**1.41**

Table 6.3: Statistics for the most preferred version.

are not entirely sure about the relation between the music and the respective preferred image. A more in-depth analysis of these results is described below, in conjunction with the adjectival description of each music-image pair. It is important to note that we describe emotions like happiness, enjoyment, and excitement as positive type of emotions, while sadness and boredom as negative.

For the least preferred image, its statistics can be seen in table 6.4. Here, by considering the first question (How would you rate this image?), and looking at the statistics of the mean, median, and mode, we observe that the results are lower in comparison with the previous preferred image - it was expected, since we are assessing the least preferred version. In the second question (Do you think this image is related to the music?), the answers are, once again, lower in comparison with the preferred image. This means that participants consider that the least preferred image has less relationship with the respective music that served as inspiration for the generation process.

Regarding the adjectival description of each image, we have two different cases: the first, where the participants had to describe the three image versions and respective music according to their perception (open answer), and the second, where a predefined list of adjectives based on basic emotions was provided to be chosen.

Regarding the first case, for the Handel - *Concerto No. 1*, the results for the music and respective preferred image can be seen in figure 6.3. Here, by looking at the results, we can see that, for both image and music, the opinions are much divided. The image was mainly perceived as happy and enjoyable, while the music was perceived as sad, smooth and fearful. These results are contradictory since they transmit opposite sensations. The positive emotions related to the image can be related to the image background, that is mainly composed of warm colors distributed in vertical stripes. Besides,

Question	Music	Mean	Median	Mode	Standard Deviation
How would you rate this image?	Handel - *Concerto No. 1* *2. Largo*	**3.35**	**4**	**4**	**1.31**
	Handel - *The Royal Fireworks Suite* *3. La Paix*	**2.95**	**3**	**3**	**0.99**
	Mozart - *Symphony No. 40* 3. *Menuetto and Trio; Allegretto*	**2.51**	**2**	**3**	**1.05**
	Stravinsky - *The Firebird* *1. Introduction*	**2.49**	**2**	**2**	**1.04**
Question	**Music**	**Mean**	**Median**	**Mode**	**Standard Deviation**
Do you think this image is related to the music?	Handel - *Concerto No. 1* *2. Largo*	**2.98**	**3**	**4**	**1.25**
	Handel - *The Royal Fireworks Suite* *3. La Paix*	**2.89**	**3**	**3**	**0.99**
	Mozart - *Symphony No. 40* 3. *Menuetto and Trio; Allegretto*	**2.89**	**3**	**3**	**1.05**
	Stravinsky - *The Firebird* *1. Introduction*	**3.31**	**4**	**4**	**1.26**

Table 6.4: Statistics for the least preferred version.

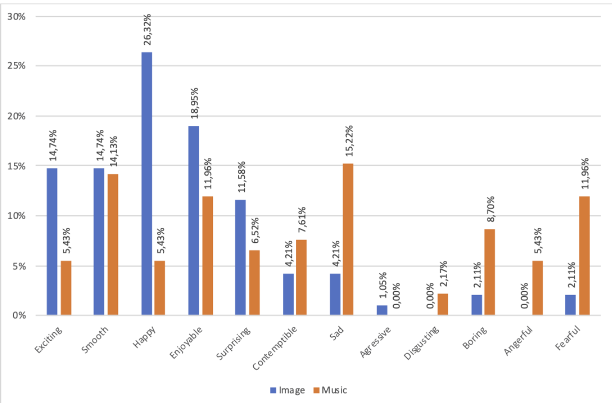


Figure 6.3: Description obtained from preferred of Handel - *Concerto No. 1*.

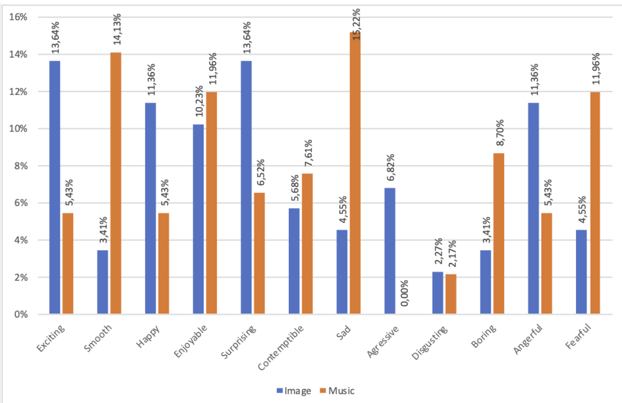


Figure 6.4: Description obtained from least preferred of Handel - *Concerto No. 1*.

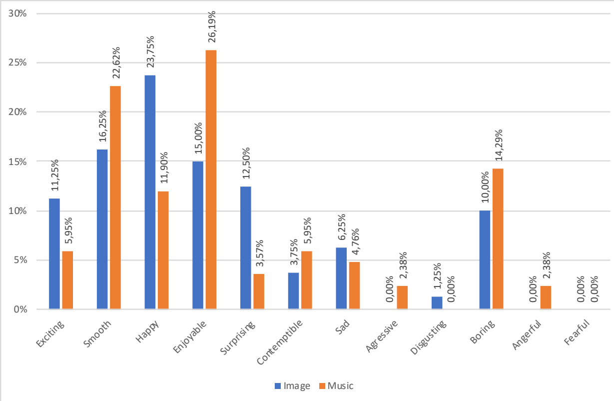


Figure 6.5: Description obtained from preferred of Handel - *The Royal Fireworks Suite*.

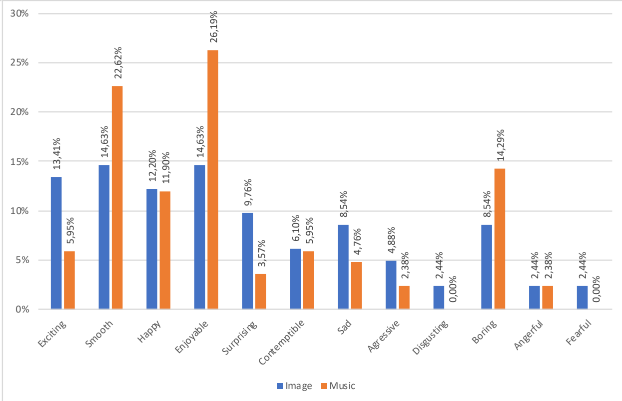


Figure 6.6: Description obtained from least preferred of Handel - *The Royal Fireworks Suite*.

this color distribution continues to the foreground, where each element has high transparency related to high-pitched elements presented in the music. In turn, the music has a low pace or rhythm, that may be related to the sadness perceived. Nonetheless, considering that the participants were describing different artworks, its reasonable to have disagreements. Concerning the open question, participants mostly described the image as abstract, colorful and geometric. Besides, they perceive many different geometric shapes that transmit the idea of movement in a 3-dimensional space. For the music, people described it as tense, nostalgic, and sad, reinforcing the adjectival description previously made. All these results can be confirmed in table 6.3, where the mean and mode for the relation between the music and image is 2, and the median 2.38. Since the results are under 3, we consider that participants do not perceive any relation between the music and image, which is a reasonable explanation for the contradictory opinions they have. For the least preferred image, the results are similar - figure 6.4. Interestingly, some participants perceive a “rain of color” in a repetitive pattern through the image. Besides, some said that all the image versions were related, and that they continue each other in the patterns. However, for the music and both most and least preferred images, the results for the “enjoyable” perception are balanced. This can be confirmed in tables 6.3 and 6.4, where the median and mode for the image rating are 4, and mean above 3.

The second case is the Handel - *The Royal Fireworks Suite*, which results are more aligned in comparison with the previous case. The answers about the musical and respective preferred image can be seen in figure 6.5. Here, by looking at the results, participants mainly perceive the image as happy and smooth, while the music as enjoyable and smooth. Although we were expecting a more subjective description of the images, participants mainly described what they observe in the canvas: lines, circles, dots and other geometrical shapes. Some answered that the image is colorful and that it can be considered an animated version of binary code. Nonetheless, the majority used descriptors that belong to the positive type of emotions. Indeed, the transparency levels of the elements that compose the images are

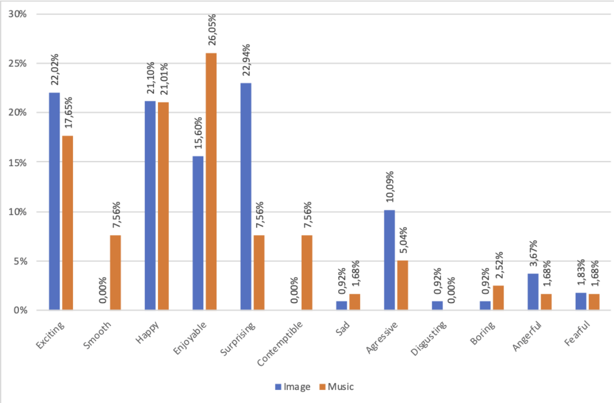


Figure 6.7: Description obtained from preferred of Mozart - *Symphony No. 40*.

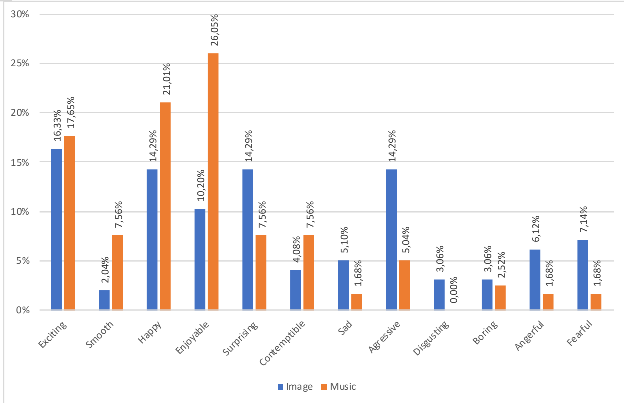


Figure 6.8: Description obtained from least preferred of Mozart - *Symphony No. 40*.

low, which leads to a better perception of geometric shapes and their boundaries, enhancing their color levels. In turn, regarding the music's description, the used descriptors are similar, and can be confirmed by the statistics on table 6.3, where the mean for the image rating is 3.7, while the median and mode are 4. Regarding the relation with the music, since the mean is 2.93, median 3, and mode 4, we consider that a better relationship could be perceived with it, since the used descriptors belong to the same type of emotions. For the least preferred image, there is no slight difference, since most chosen descriptors were the same as with the preferred image - figure 6.6. Regarding the participants' own perception, some keep saying that the image looks like the continuation of the previous, while others said that it is monotonous, aligned with patterns. These opinions may be related to the image's greenish tones, as well as with the vertical arrangement of the elements. Once we are assessing the least preferred version, this similarity on the descriptors was not expected. However, by looking at table 6.4, for both questions, we observe that the mode and median are 3, while the mean is lower in comparison with the previous version. This means that participants enjoyed both the images, but perceive less relation between the music and the least preferred image. Anyway, these results are positive, since the peaks for all the descriptors mainly match.

The third case is the Mozart - *Symphony No. 40*, whose descriptors for the preferred image and respective music can be seen in figure 6.7, while the descriptors for the least preferred in figure 6.8. Here, participants mainly perceive the preferred image as surprising and exciting, while the music as enjoyable and happy. Regarding their own perceptions, likewise the first case, some people said that the image is abstract, colorful and geometric. We believe that these image descriptors were used because participants perceived the considerable increase of elements in the image, that may transmit the sensation of movement. In turn, some people described the music as harmonic and beautiful. Despite the wide range of used vocabulary, we consider that the descriptions for both the artifacts remain to the same emotional categorization, which leads to an agreement between descriptors. This relation

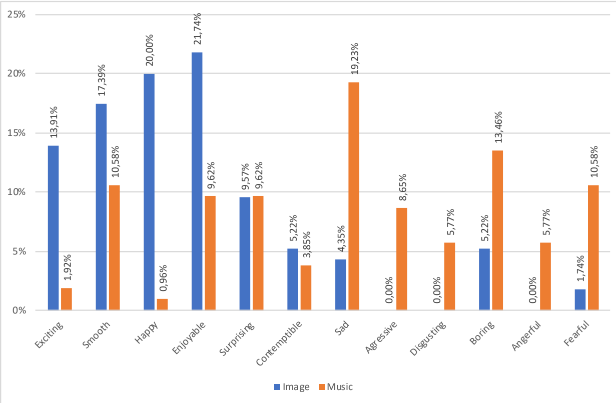


Figure 6.9: Description obtained from preferred of Stravinsky - *The Firebird*.

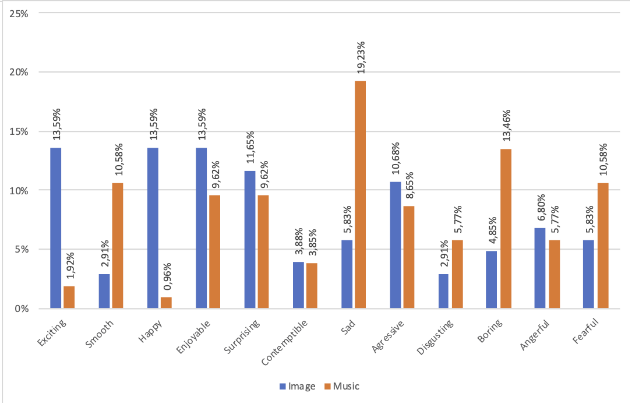


Figure 6.10: Description obtained from least preferred of Stravinsky - *The Firebird*.

can be confirmed in table 6.3, since the mean for the relationship between both artifacts is 3.23, while the median 3, and mode 4. For the image rating, the results keep positive, since the mean is 3.68, while the median and mode 4. This means that most of the participants enjoyed the image and did perceive a relation between both artifacts. For the least preferred image, despite the peak on the exciting descriptor, participants also consider the image to be happy, surprising, and aggressive. Its statistics are in table 6.4, where we observe the lower values for the image rating. In the end, some said that the image resembles a stage full of bodies, or a roller coaster full of people. We believe that these opinions may be related to the canvas's arrangement of various shapes, due to the diversity of instruments that compose the symphony.

The fourth and last case is the Stravinsky - *The Firebird*, whose descriptors for the preferred image and respective music can be seen in figure 6.9. Here, as in the first case, by observing the results we see that the opinions are contradictory for the music-image pair - participants perceived the music mainly as sad and boring, while the preferred image as enjoyable and happy. Regarding the image description, people perceived a “rain of colors”, or dropping colors. Besides, it is interesting to note that this particular image was described as artificial, probably made by a computer, where through lines of inspiration, circles, and hearts transmit the idea of color progression. We consider that these descriptions are related to the spatial distribution of different colored elements over the canvas, that is directly related to the various rhythmic values throughout the music. In turn, the music was described mainly as sad and depressing, related to suspense soundtracks. Indeed, this music has more lower-pitched elements, in a slow-paced rhythm that directly influence the lightness of each visual element. However, since this image is composed of darker elements, we were not expecting descriptors that would relate to the positive type of emotions. Considering the statistics regarding the relationship between both artifacts - table 6.3 - we observe that the mode is 2, the lowest of the four cases - this explains the opposite perceptions participants had. However, the statistics for the image rating have the most favourable

Question	Music	No	Yes	Total
Do you consider these to be abstract images?	Handel - *Concerto No. 1* *2. Largo*	**17.50%**	**82.50%**	**100%**
	Handel - *The Royal Fireworks Suite* *3. La Paix*	**18.18%**	**81.82%**	**100%**
	Mozart - *Symphony No. 40* *3. Menuetto and Trio; Allegretto*	**13.21%**	**86.79%**	**100%**
	Stravinsky - *The Firebird* *1. Introduction*	**14.29%**	**85.71%**	**100%**
	Total	**14.29%**	**84.41%**	**100%**

Table 6.5: Do you consider these to be abstract images?

results, since the mean for the preferred image is 3.16, and mode 5. Regarding the least preferred image, the results can be seen in figure 6.10. Here, for the same music's descriptors, the least preferred image was perceived equally exciting, happy and enjoyable. However, its statistic values - table 6.4 - for the image rating are lower than the preferred version, what was expected. Nonetheless, in comparison with the preferred image, participants consider there is more relation between the image and music. Finally, considering the description participants did, we observed that some described the images as enjoyable, but sometimes boring, transmitting two contradictory emotions. This can be explained by the fact that people may like the artifacts since they convey different visualizations with the most various shapes and colors. At the same time, we believe that the boring descriptor was chosen because, for all the generated images, each version was equally generated, i.e., with the same colors, positions and other features, that may have ended up being similar.

We cannot fail to point out that, despite the individual analysis, all the images were generally described with adjectives that belong to positive type of emotions, such as happiness, excitement, or joy. We also noticed that participants tried both to explain what they observed, combined with the aroused emotions. The majority considered that images are colorful, geometric, somehow following one pattern. In some cases, people said that had a sensation of movement through the images, and that somehow they were a combination of each other. Indeed, since all the images were inspired on music, considering that music is a combination of sounds (and silence) throughout time, following rules that describe its structure, it is reasonable that images were perceived this way. Finally, having in mind that each survey evaluates two sets of music-image, we believe that participants could be biased to the second set, by the first list of predefined adjectives. Therefore, by comparing all the answers on different artifacts, we did a detailed analysis of each survey to verify if the description people wrote about the artifacts was biased. We concluded that, despite the repetition of some descriptors, the only difference is that the quantity of sentences used to describe the images decreases through the survey, which is a sign of tiredness. It is acceptable to have this reduction on the amount of descriptions since we consider the survey to be dense and repetitive.

Having finished the adjectival description of the most and least preferred versions, we can move on to the question - Do you consider these to be abstract images? - that was applied to all the generated set of images. The relative frequency for each music and respective image set can be seen in table 6.5. The results were positive since the majority of the participants (around 84.71%) consider the images to be abstract. Besides, “abstract” was one of the most used concepts to describe the images through all the versions.

Regarding the last survey's question - ”All the images were generated having the previous music as inspiration” - Do you agree with this statement? - the results are still positive and in line with our expectations and goals - table 6.6. Here, by observing the statistics, we can see that the majority of the participants believe that there is a relation between the music and image - the median and mode for the first and second set of images are 3, while for the third and fourth set, the value is 4 in both cases. In turn, the mean is around 3, that is still positive.

A complete analysis considering different factors - gender, age group, or the frequency participants usually go to exhibitions - was also made. However, it is not presented, given the little variability of the statistics. We noticed that the main difference is that people aged between $50 - 59$ performed a more subjective description of the artifacts, while the majority, aged between $18 - 29$ described them with fewer, and more objective concepts, based on what they objectively saw in the images. Since these do not bias the results regarding our objectives, we decided to present the results with no filters on these factors.

Since our main objectives were to generate abstract images that have as inspiration musical files, we consider that the results are favorable to our objectives. From the surveys, we obtained answers to verify the goals related to the image's quality and the association between each image and music. First, most of the participants consider that the generated images are abstract. Second, regarding the image-music association, we believe that the results were also good, since the average from the relation between the artifacts is positive. Besides, in two of the four studied cases, the peaks on the used descriptors mostly match, being reinforced by the written descriptions, which means that both the artifacts were perceived the same way. Finally, even though it was not our main objective, regarding the preferred images, the ratings are good - generally, participants feel to like the artifacts, as already seen. Regarding the least preferred images, the results are more neutral, but still flattering. From these, we can state that participants liked all the images, even the least preferred ones.

In the end, we generated three different images considering different shapes' associations, and recalling that the most preferred version was the Genetic, we believe that our algorithm reached the goals. Regarding the creativity domain, our conclusions are much more subjective. Recalling the definition of creativity presented in Section 2.1, creative artifacts shall be novel and useful. From the results, we observe that participants generally rated our images with a 3 or higher, which means that they enjoyed

Question	Music	Mean	Median	Mode	Standard Deviation
Do you agree with the previous statement?	Handel - *Concerto No. 1* *2. Largo*	**2.95**	**3**	**3**	**1.11**
	Handel - *The Royal Fireworks Suite* *3. La Paix*	**3.11**	**3**	**3**	**1.13**
	Mozart - *Symphony No. 40* *3. Menuetto and Trio; Allegretto*	**3.43**	**4**	**4**	**1.07**
	Stravinsky - *The Firebird* *1. Introduction*	**3.45**	**4**	**4**	**1.16**

Table 6.6: "All the images were generated having the previous music as inspiration" - Do you agree with this statement?

the generated artifacts. However, as previously stated, whether people like or dislike something has to do with temporal, cultural, motivational and emotional factors (considerations that machines have almost nothing to say).

7

Conclusion and Future Work

Creativity remains to be an interdisciplinary ability that is essential when creating something new, or solving problems. It is directly linked with inspiration, that may come from our own subconscious, or from the surrounding environment, depending on the culture, time of the history, or if we are doing something individually or in a group, gathering cultural, social, and personal forces. In this Thesis, we studied definitions and theories that emerged throughout the history about creativity, and concluded that it is far from being understood and thoroughly explained. Similarly to creativity, inspiration has no widely accepted definition about it, but these concepts seem closely related. There is no established theory that explains it, neither a clear explanation about how it works in human brains.

In our work, we focus on the Computational Creativity area, a sub-field of Artificial Intelligence research that studies how software can exhibit behaviour that unbiased observers would deem to be creative. Since creativity and inspiration are two related concepts, regarding Computational Creativity, an association or one mapping between two domains shall be established to create a system that draws creativity inspired on other artifacts. Therefore, we explained some inspirational systems, and the most used techniques to reach their goals.

Motivated by the algorithmic way of processing both music and image, the main objective of our Thesis was to develop a system that exhibits creative behaviour through inter-domain associations, by generating abstract images inspired on musical artifacts. Our approach was based on the relation between the music's harmony, melody and rhythm, and visual elements that refer to both image's background and foreground. Through trial and error, we chose to relate the music's melody to the foreground, music's harmony to the background, and music's rhythm to the size of the generated elements. It is important to note that one correct association between two domains does not exist, since the developer's choice directly influences the analogies made.

First, we did extensive background research on music theory, so that we could understand the domain in which our system draws inspiration from. MIDI files were used to better understand the music's basic elements, how they are presented, structured, and organized. With this, and considering our approach, we obtained enough information regarding which features to use. Second, we researched on color harmony, shape assembling, and image generation techniques, so we could better understand how to associate all the features that come from the music given as input. The associations that we propose do not impose any restriction regarding the image generation. However, in the musical domain, we decided to restrict the input files to music that last between three and five minutes. Besides, due to its diversity and composition of sophisticated forms of instrumental music, we restricted the style to Classical music, although other styles could be used. Finally, default values were chosen to the dimensions of the image - 1600$\times$2700 pixels - even though other dimension values can be defined by the users at the beginning of the run of the system. Besides, there is no limitation regarding the screens where the images are displayed.

Three images were generated from each music file given as input. The first, called the **Random** Image, was generated assigning a random shape and texture to each instrument found in the music. The second, called the **Associated** image, was generated assigning a predefined association between instruments and respective shapes. This predefined association was entirely based on our personal perception, since an association between instruments and shapes that has a reasonable explanation is hard to achieve. It is also important to note that a study on these associations was made and deeply tested through an online platform developed for this goal. With this, users can test their different ideas by creating their own inter-domain associations. The third, called the **Genetic** Image, results from the execution of a Genetic Algorithm that receives the two previous versions to generate the initial population. Their evolution considers color and music harmonies between nearest elements in the generated images, as well as visual perceptions of color, location of the elements in the canvas, and shape's distributions.

We evaluated four different music files and respective generated images distributed in four different online surveys. The provided musics are from different periods of time, and different composers. The majority of the participants considered that the generated images are abstract, and do believe that they have a relation with the music that served as the basis for the inspiration process. Besides, the majority liked the presented images, ranking the genetic as the most preferred one. We consider that these results are favourable since they go in line with our goals. A special mention for the positive feedback we had in most of the generated images. Finally, the most written description by participants was "abstract", reinforcing what we were expecting.

The system we developed has some limitations, especially regarding the dimensions of each element in the canvas, and the color levels fluctuation among different screens. Therefore, we tried to overcome these by applying two filters to all the generated images, as well as to give the user's the possibility to parameterize the variables related to the dimensions of the image. As future work, we consider that small changes can be made to our approach regarding the associations made between instruments and shapes, so that a better reflection of the essence of each timbre could be perceived. Besides, if MP3 files were used instead of MIDI files, by using signal processing techniques, one can study the harmonic series of each instrument found on the musical sheet and build a dedicated shape for each instrument. Moreover, the instrument's timbre found on MIDI files is very poor, and possibly some essence of the music can be lost in the association process. However, despite all these issues, we consider that the presented approach, among many that could have been made, reaches our goals. Nevertheless, we hope that a full music signal analysis approach to generate visual art can be considered as future work, as well as the generation of video animations, so that the visual elements can go along with the music over the time.

In the end, engineering software for creative proposes has been introducing a new era where machines can be seen as a tool for artists in their creative process, as well as independent and creative agents in the most diverse domains, contributing thus in interesting ways to human culture. We believe that these inspirational systems that rely on cross-domain associations enable two-way communication between them and their users, either to explain processes and decisions, support ideas and learn from this collaboration to improve the creative process. Purists in respect to human uniqueness say that the artifacts produced by machines are not considered art, thus the software can not be considered creative by itself - only humans can make it [34]. With the development of our system, we believe that this may not be always true, since our system can generate abstract images having inspiration on other artifacts - this explains why it may also be recognized that our system has, at least, some degree of creativity. Nonetheless, with this work, we developed a system that generates abstract images inspired from music, thus contributing to the research made in the Computational Creativity area, especially in the field of image generation.

Bibliography

[1] Charles Ames and Michael Domino. Cybernetic Composer: an overview. In *Understanding music with AI*, pages 186–205. MIT Press, Cambridge, MA, 1992.

[2] Kim Binsted. Machine humour: An implemented model of puns. Master's thesis, The University of Edinburgh: College of Science and Engineering, 1996.

[3] Margaret A Boden. Creativity and Artificial Intelligence. *Artificial Intelligence*, 103(1-2):347–356, 1998.

[4] Margaret A Boden. Computer models of creativity. *AI Magazine*, 30(3):23–23, 2009.

[5] Francisco Braga, H. Sofia Pinto, and Luís Silveira. Sculpture Inspired Musical Composition : One Possible Approach. Master's thesis, Universidade de Lisboa: Instituto Superior Técnico, 2019.

[6] Simon Colton. The Painting Fool - A Computer Artist. http://www.thepaintingfool.com/, Accessed: 2020-04-12.

[7] Simon Colton. Creativity Versus the Perception of Creativity in Computational Systems. In *Proceedings of the AAAI spring symposium: creative intelligent systems*, volume 8, 2008.

[8] Simon Colton. The Painting Fool: Stories from Building an Automated Painter. In *Computers and Creativity*, pages 3–38. 2012.

[9] Computational Creativity. http://computationalcreativity.net/home/about/computational-creativity/, Accessed: 2020-03-18.

[10] David Cope. Recombinant music: using the computer to explore musical style. *Computer*, 24(7):22–28, 1991.

[11] Steve DiPaola and Liane Gabora. Incorporating characteristics of human creativity into an evolutionary art algorithm. *Genetic Programming and Evolvable Machines*, 10(2):97–110, 2008.

[12] Kemal Ebcioğlu. An expert system for harmonizing four-part chorales. *Computer Music Journal*, 12(3):43–51, 1988.

[13] Paul Ekman. Universal Emotions. https://www.paulekman.com/universal-emotions/, Accessed: 2020-01-02.

[14] Ahmed Elgammal, Bingchen Liu, Mohamed Elhoseiny, and Marian Mazzone. CAN: Creative Adversarial Networks, Generating "Art" by Learning About Styles and Deviating from Style Norms. *arXiv preprint arXiv:1706.07068*, 2017.

[15] Shelley Esaak. The 7 Elements of Art and Why Knowing Them Is Important. https://www.thoughtco.com/what-are-the-elements-of-art-182704, Accessed: 2020-07-16.

[16] Shelley Esaak. The Definition of Shape in Art. https://www.thoughtco.com/definition-of-shape-in-art-182463, Accessed: 2020-07-16.

[17] Shelley Esaak. What Is the Definition of Color in Art? http://thoughtco.com/definition-of-color-in-art-182429, Accessed: 2020-07-13.

[18] ETC Video Synthesizer. https://www.critterandguitari.com/etc, Accessed: 2019-11-10.

[19] Philip Galanter. Computational aesthetic evaluation: past and future. In *Computers and Creativity*, pages 255–293. Springer, 2012.

[20] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. In *Advances in neural information processing systems*, pages 2672–2680, 2014.

[21] Douglas R Hofstadter, Melanie Mitchell, et al. The copycat project: A model of mental fluidity and analogy-making. *Advances in connectionist and neural computation theory*, 2:205–267, 1995.

[22] Britton Horn, Gillian Smith, Rania Masri, and Janos Stone. Visual information vases: Towards a framework for transmedia creative inspiration. In *ICCC*, pages 182–188, 2015.

[23] Anil K Jain, Jianchang Mao, and K Moidin Mohiuddin. Artificial neural networks: A tutorial. *Computer*, 29(3):31–44, 1996.

[24] James C Kaufman and Robert J Sternberg. *The Cambridge handbook of creativity*. Cambridge University Press, 2010.

[25] Kamaljeet Kaur and Ms Manpreet Kaur. Case Study of Color Model of Image Processing. *International Journal of Computer Engineering and Technology*, 6(12):65–71, 2015.

[26] Arthur Koestler. *The act of creation*. Hutchinson, 1964.

[27] Pat Langley, Herbert A Simon, Gary L Bradshaw, and Jan M Zytkow. *Scientific discovery: Computational explorations of the creative processes*. MIT press, 1987.

[28] Douglas B Lenat. The role of heuristics in learning by discovery: Three case studies. In *Machine learning*, pages 243–306. 1983.

[29] John McCall. Genetic algorithms for modelling and optimisation. *Journal of computational and Applied Mathematics*, 184(1):205–222, 2005.

[30] Jon McCormack. Facing the Future: Evolutionary Possibilities for Human-Machine Creativity. In *Natural Computing Series*, pages 417–451. 2008.

[31] Jon McCormack and Oliver Bown. Life's what you make: Niche construction and evolutionary art. In *Workshops on Applications of Evolutionary Computation*, pages 528–537. Springer, 2009.

[32] Julian F. Miller. Cartesian Genetic Programming. http://cs.ijs.si/ppsn2014/files/slides/ppsn2014-tutorial3-miller.pdf, Accessed: 2020-09-01.

[33] Jill Morton. Basic color theory. https://www.colormatters.com/color-and-design/basic-color-theory, Accessed: 2020-08-01.

[34] Leonel Moura. Robot Art: An Interview with Leonel Moura. 7(3):28–28, 2018.

[35] Music2Art: Automatic Generation of AI Art Inspired by Music. https://medium.com/@ahmed_elgammal/music2art-automatic-generation-of-ai-art-inpired-by-music-1a4c1e44f17a, Accessed: 2020-09-02.

[36] Marek Obitko. Introduction to Genetic Algorithms: Recommendations. https://www.obitko.com/tutorials/genetic-algorithms/recommendations.php, Accessed: 2020-08-01.

[37] George Papadopoulos and Geraint Wiggins. Ai methods for algorithmic composition: A survey, a critical view and future prospects. In *AISB Symposium on Musical Creativity*, volume 124, pages 110–117, 1999.

[38] Marcus Pearce. Notes on The Act of Creation by Arthur Koestler. *Pridobljeno*, 18(12):2013–2013, 2001.

[39] Rory Pq. Basic music theory for beginners - the complete guide. https://www.paulekman.com/universal-emotions/, Accessed: 2020-08-24.

[40] Mel Rhodes. An analysis of creativity. *The Phi Delta Kappan*, 42(7):305–310, 1961.

[41] Jean-Claude Risset and David L Wessel. Exploration of timbre by analysis and synthesis. *The psychology of music*, 2(151):26–58, 1982.

[42] Graeme Ritchie. Some empirical criteria for attributing creativity to a computer program. *Minds and Machines*, 17(1):67–99, 2007.

[43] Stuart Russel, Peter Norvig, et al. *Artificial Intelligence: A Modern Approach*. Pearson Education Limited, 2013.

[44] André Santos, H. Sofia Pinto, and Luís Silveira. Music Synthesis From Images. Master's thesis, Universidade de Lisboa: Instituto Superior Técnico, 2020.

[45] R Keith Sawyer. *Explaining creativity: The science of human innovation*. Oxford university press, 2011.

[46] Joana Teixeira and H. Sofia Pinto. Cross-Domain Analogy: From Image to Music. In *Proceedings of the 5th International Workshop on Musical Metacreation*, 2017.

[47] Joana Teixeira and H. Sofia Pinto. Cross-Domain Analogy: From Image to Music. Master's thesis, Universidade de Lisboa: Instituto Superior Técnico, 2019.

[48] Barry J Wythoff. Backpropagation neural networks: a tutorial. *Chemometrics and Intelligent Laboratory Systems*, 18(2):115–155, 1993.

[49] Han Zhang, Tao Xu, Hongsheng Li, Shaoting Zhang, Xiaogang Wang, Xiaolei Huang, and Dimitris N Metaxas. Stackgan: Text to photo-realistic image synthesis with stacked generative adversarial networks. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 5907–5915, 2017.

A

Generated Images

In this appendix we present the different images used in the evaluation process. Their names are related to the music used in the association process.

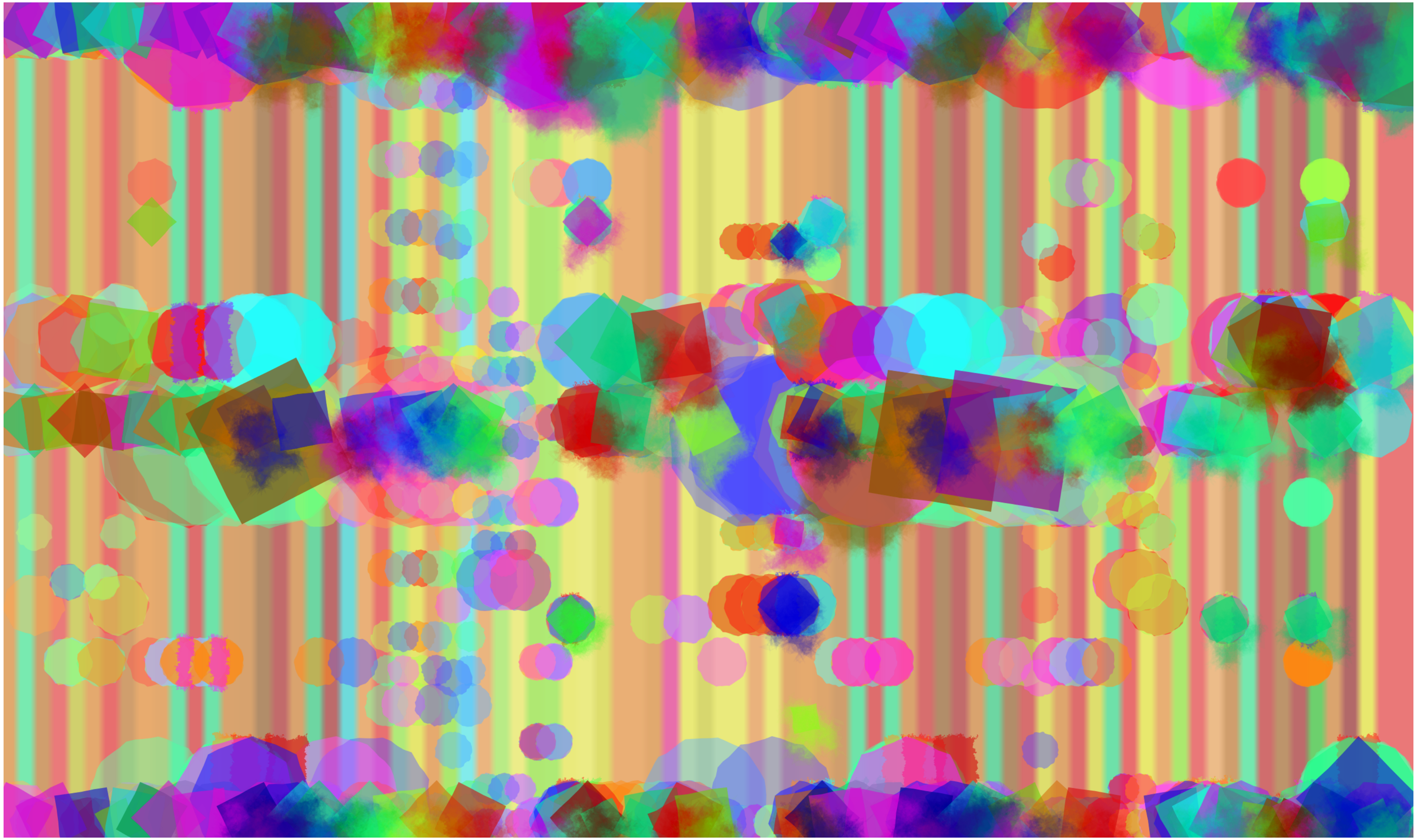


Figure A.1: Handel - *Concerto No. 1*: Random Image.

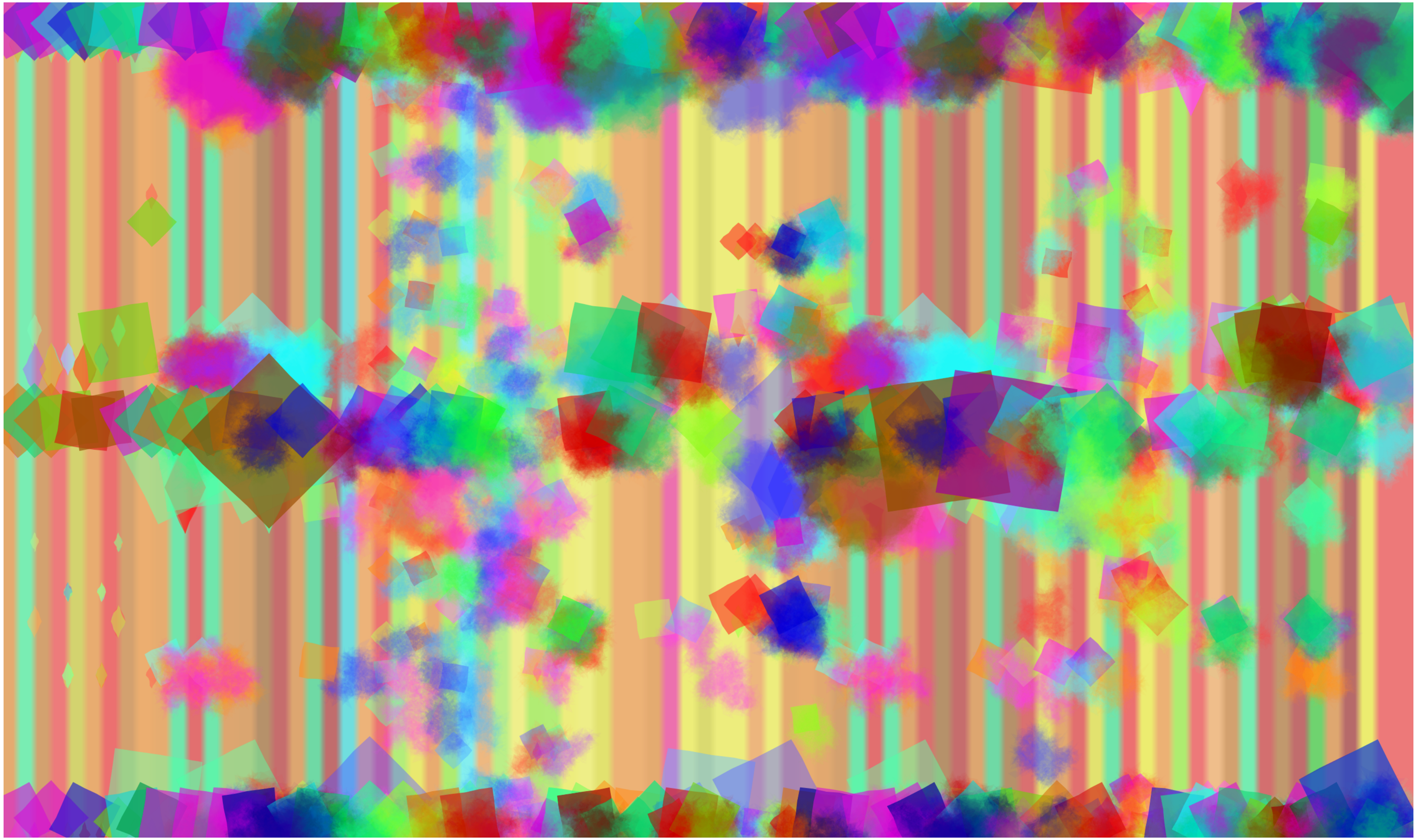


Figure A.2: Handel - *Concerto No. 1*: Associated Image.

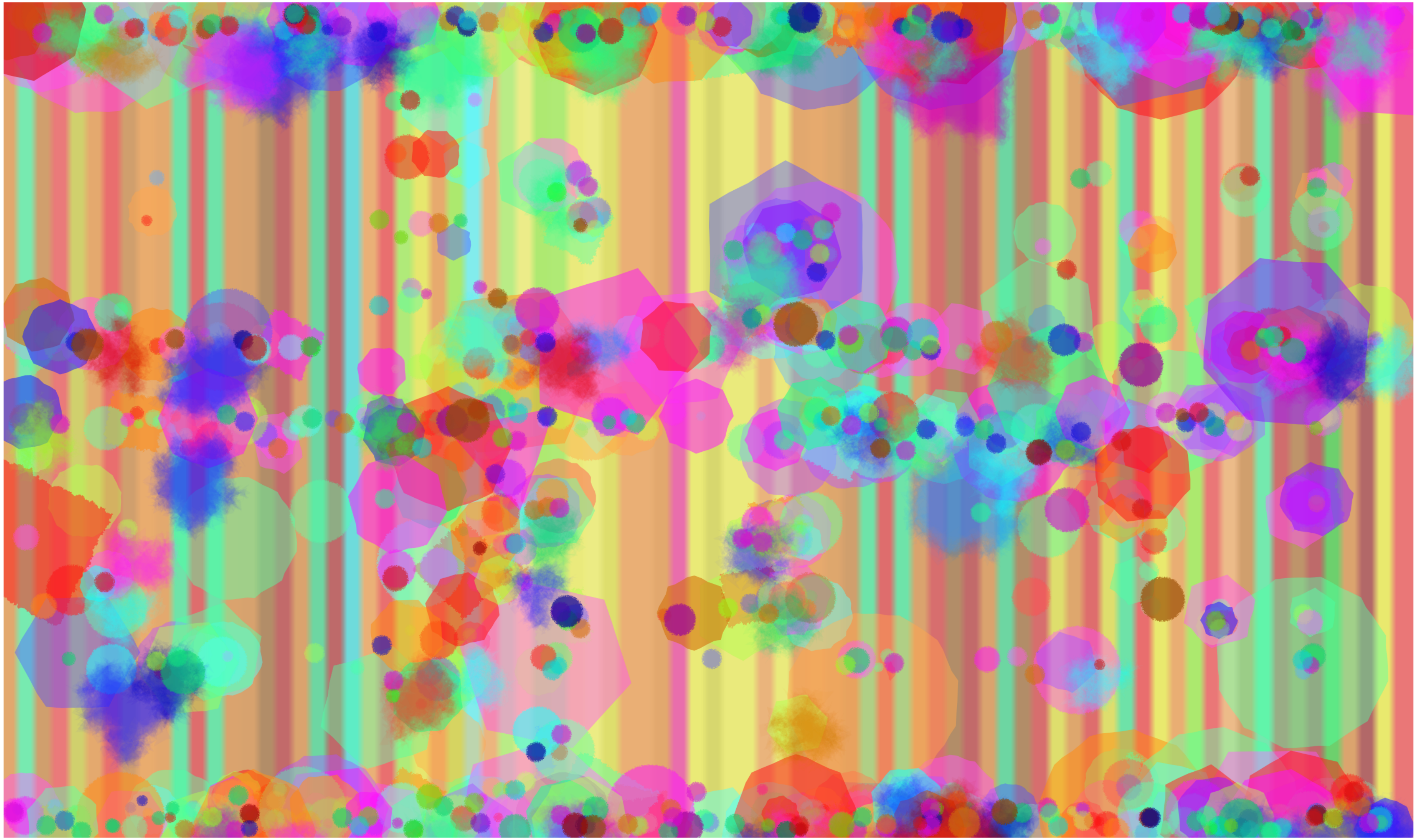


Figure A.3: Handel - *Concerto No. 1*: Genetic Image.

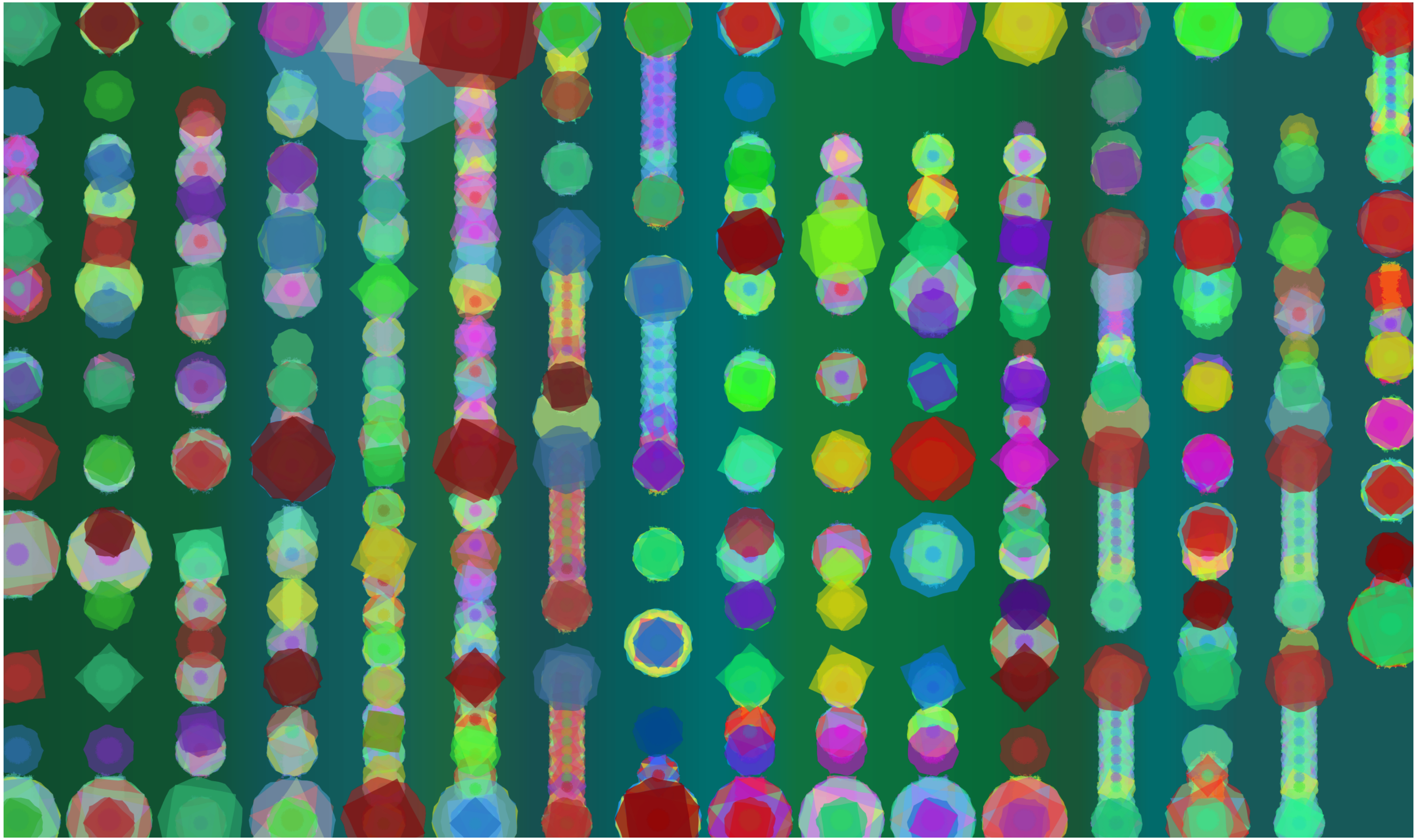


Figure A.4: Handel - *The Royal Fireworks Suite*: Random Image.

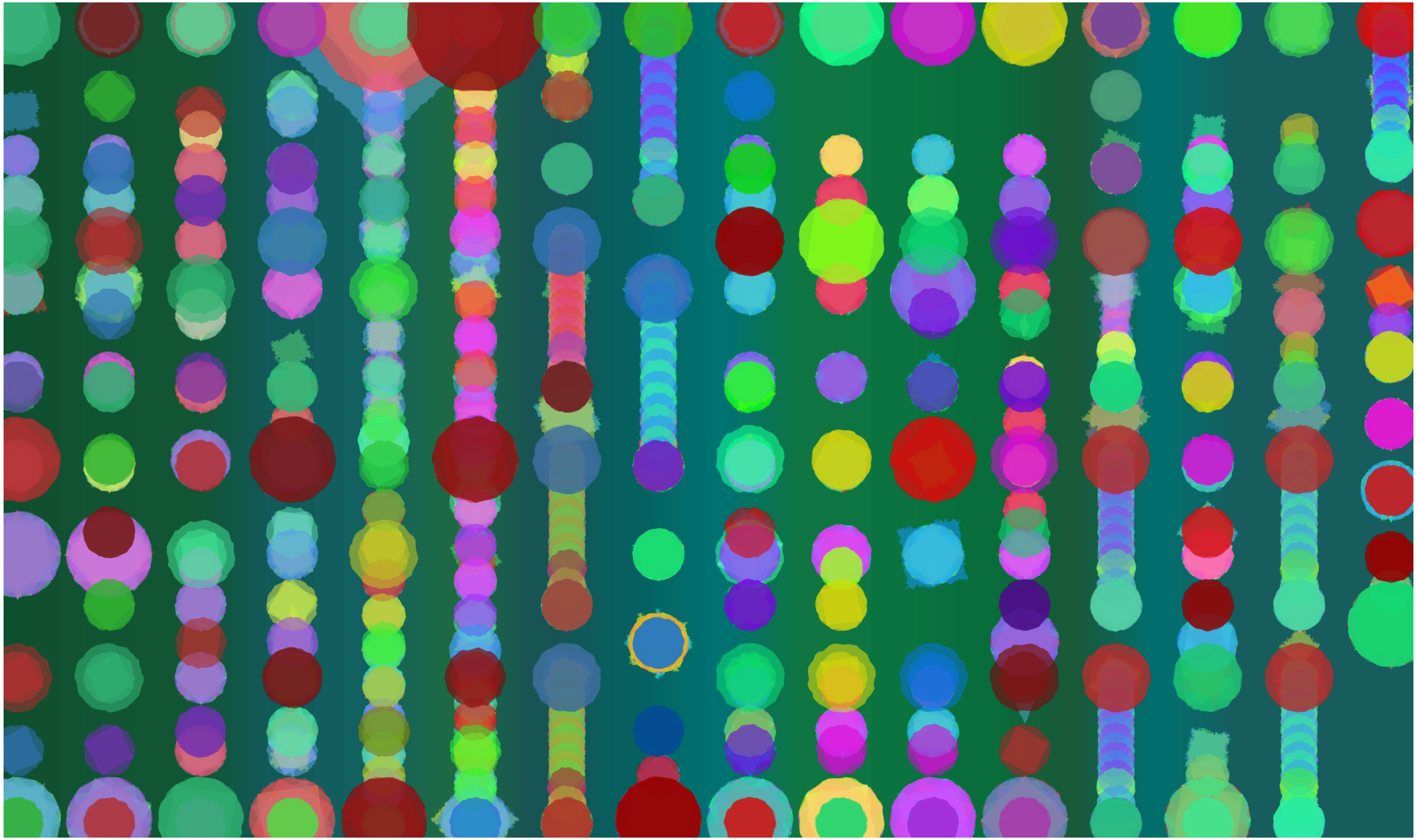


Figure A.5: Handel - *The Royal Fireworks Suite*: Associated Image.

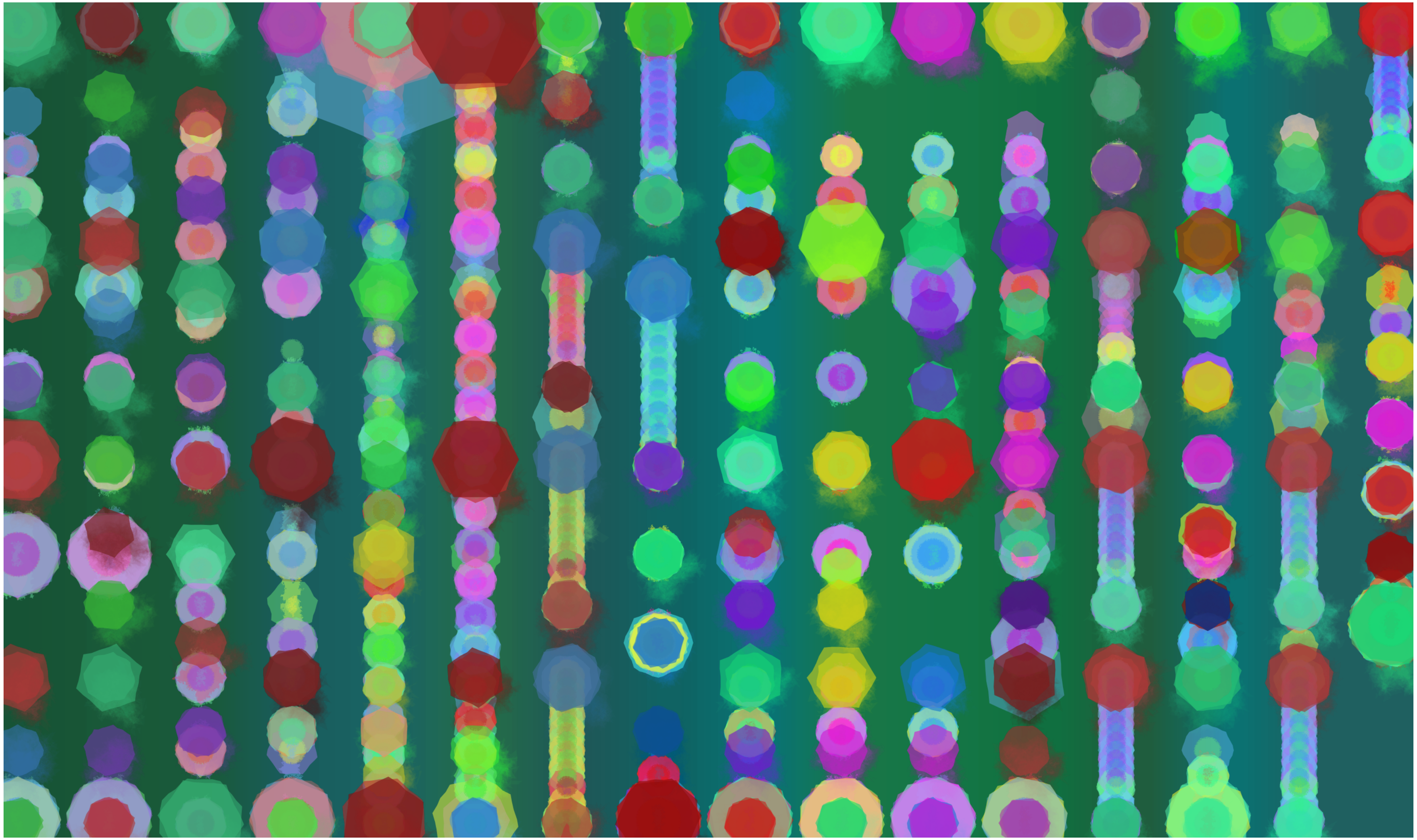


Figure A.6: Handel - *The Royal Fireworks Suite*: Genetic Image.

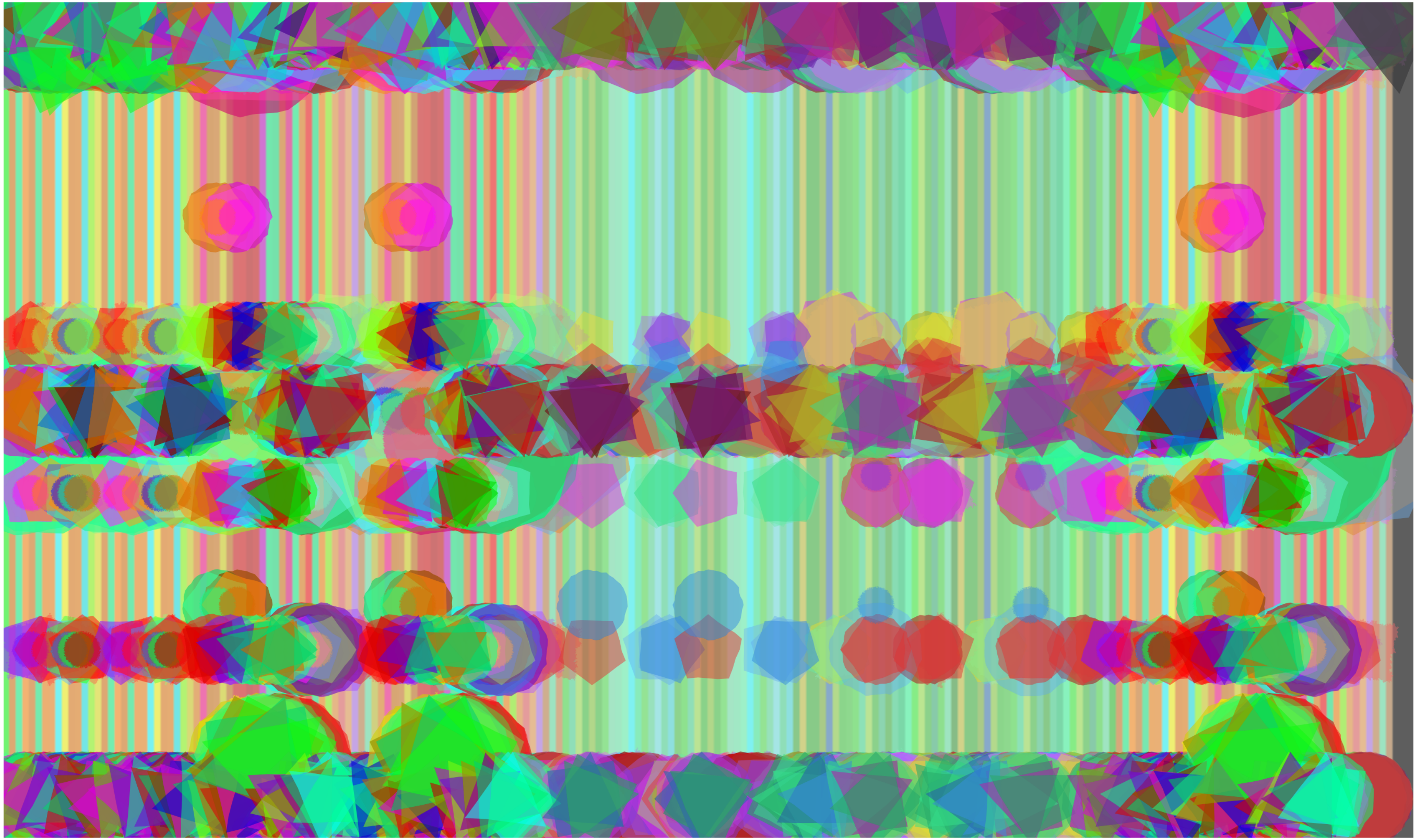


Figure A.7: Mozart - *Symphony No. 40*: Random Image.

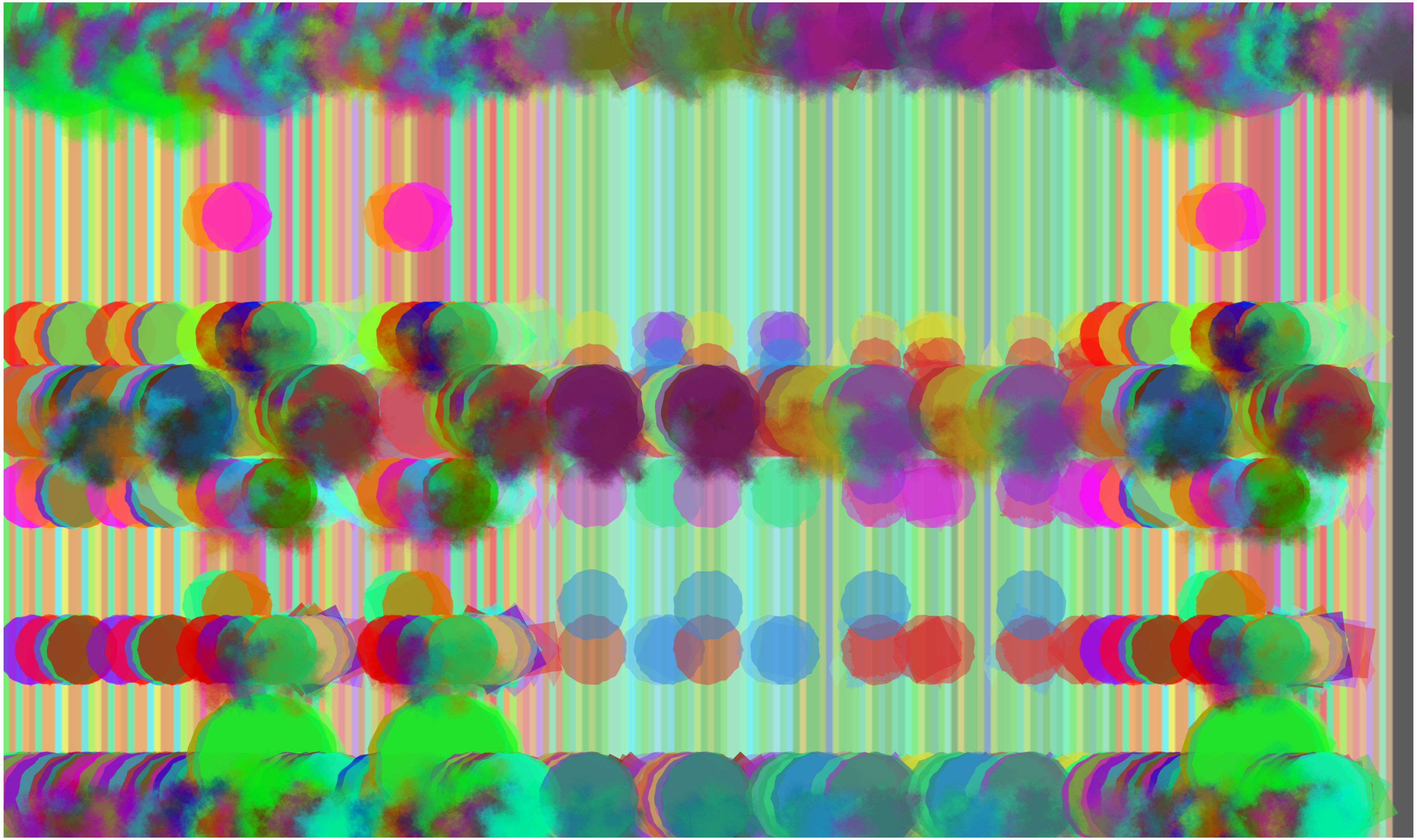


Figure A.8: Mozart - *Symphony No. 40*: Associated Image.

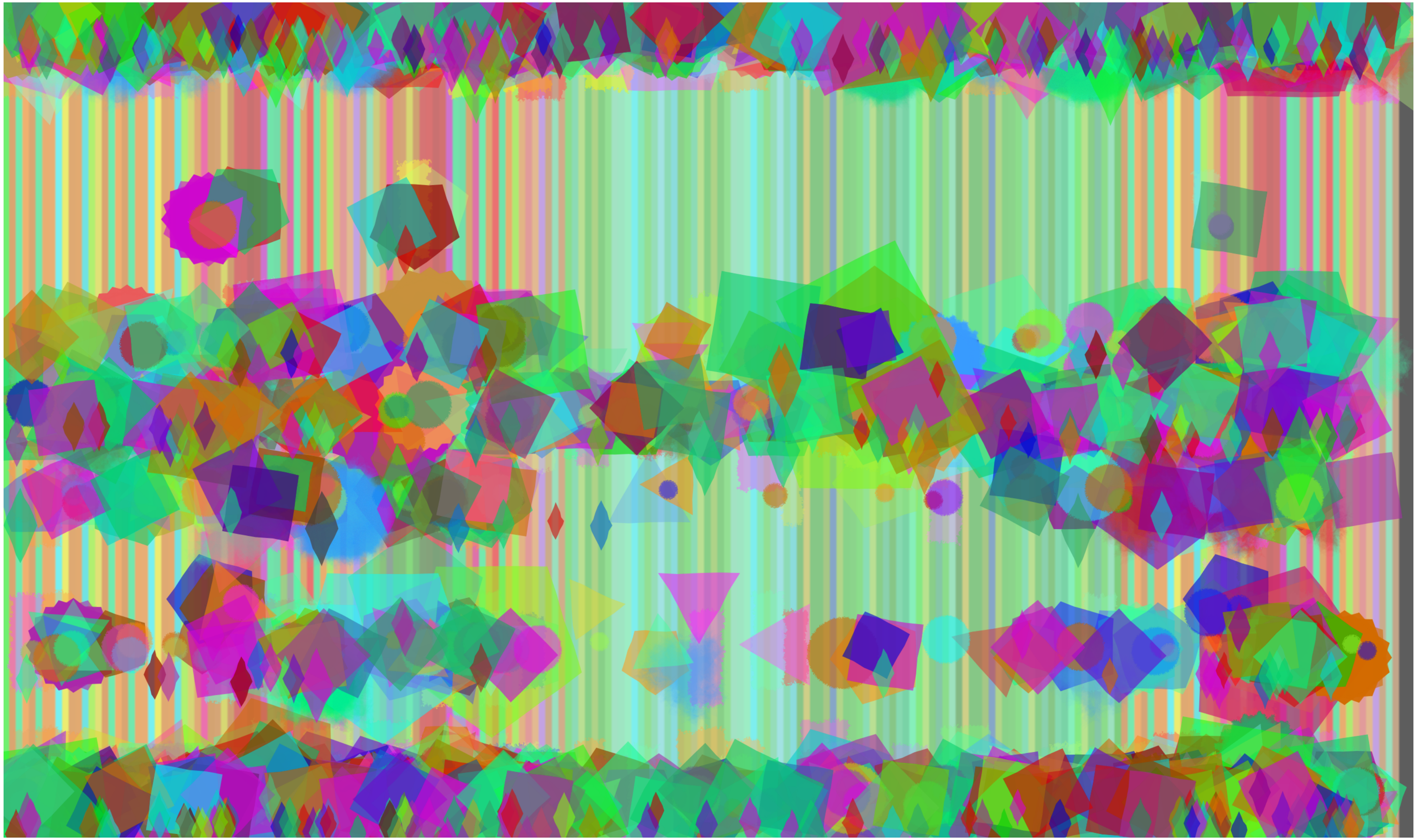


Figure A.9: Mozart - *Symphony No. 40*: Genetic Image.

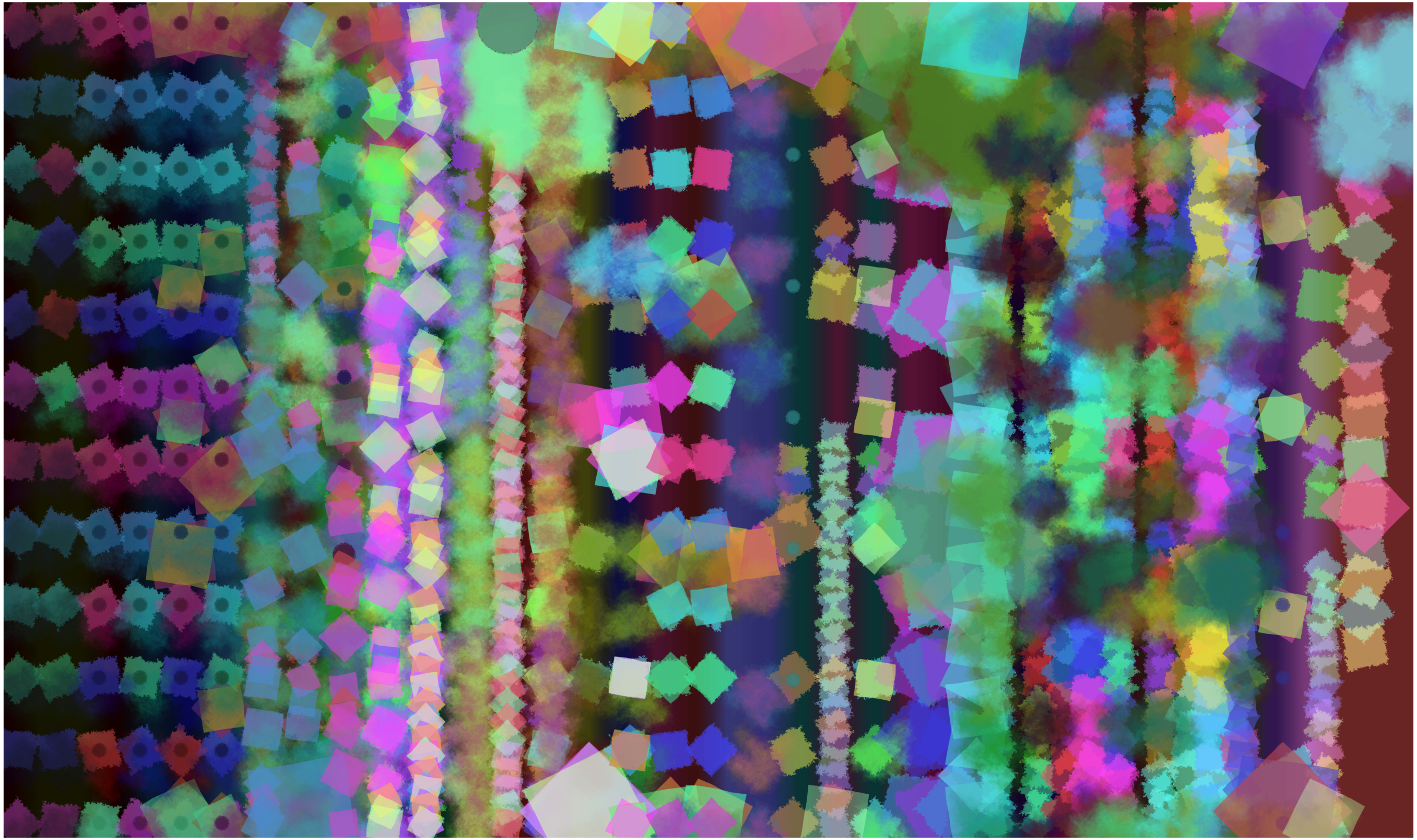


Figure A.10: Stravinsky - *The Firebird*: Random Image.

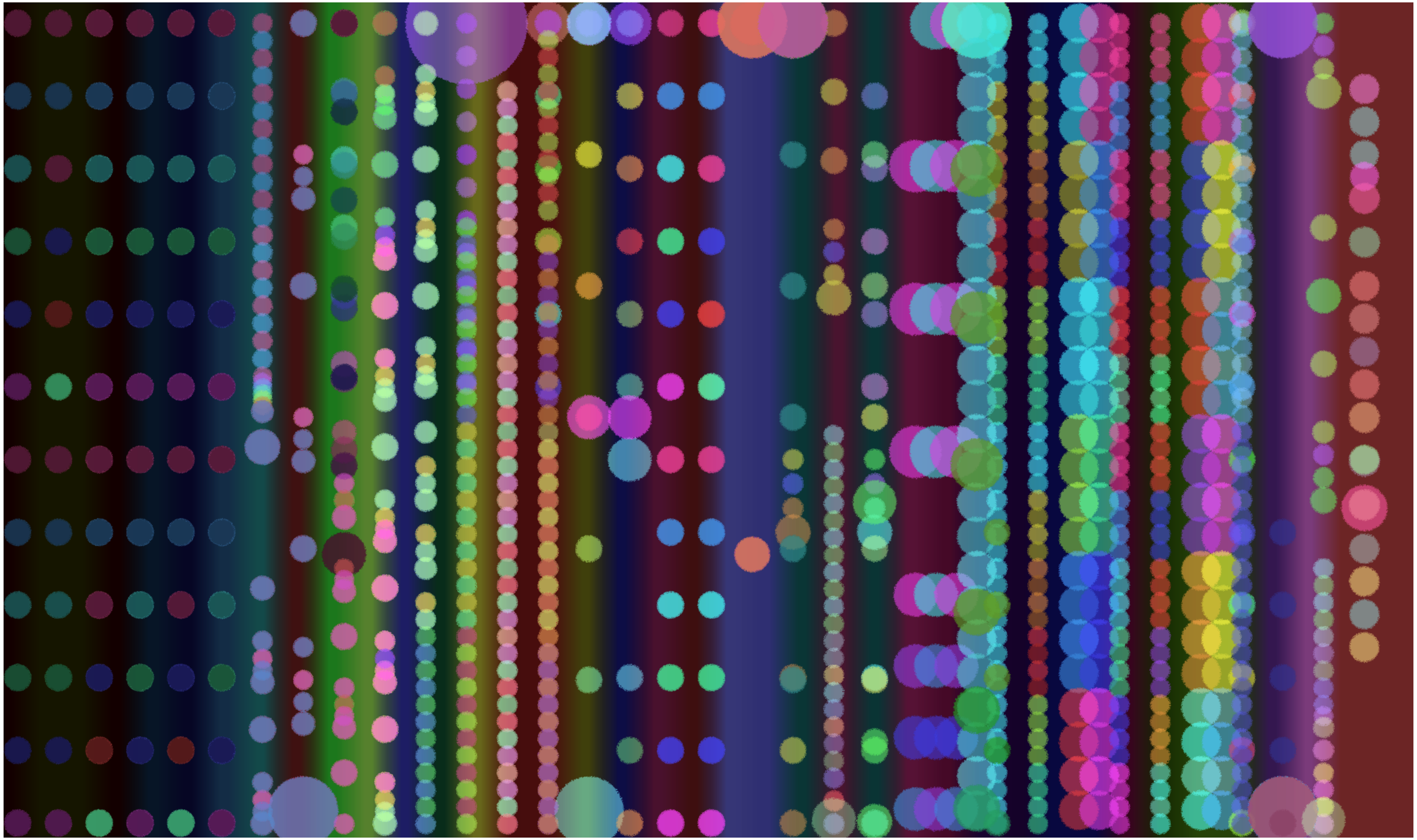


Figure A.11: Stravinsky - *The Firebird*: Associated Image.

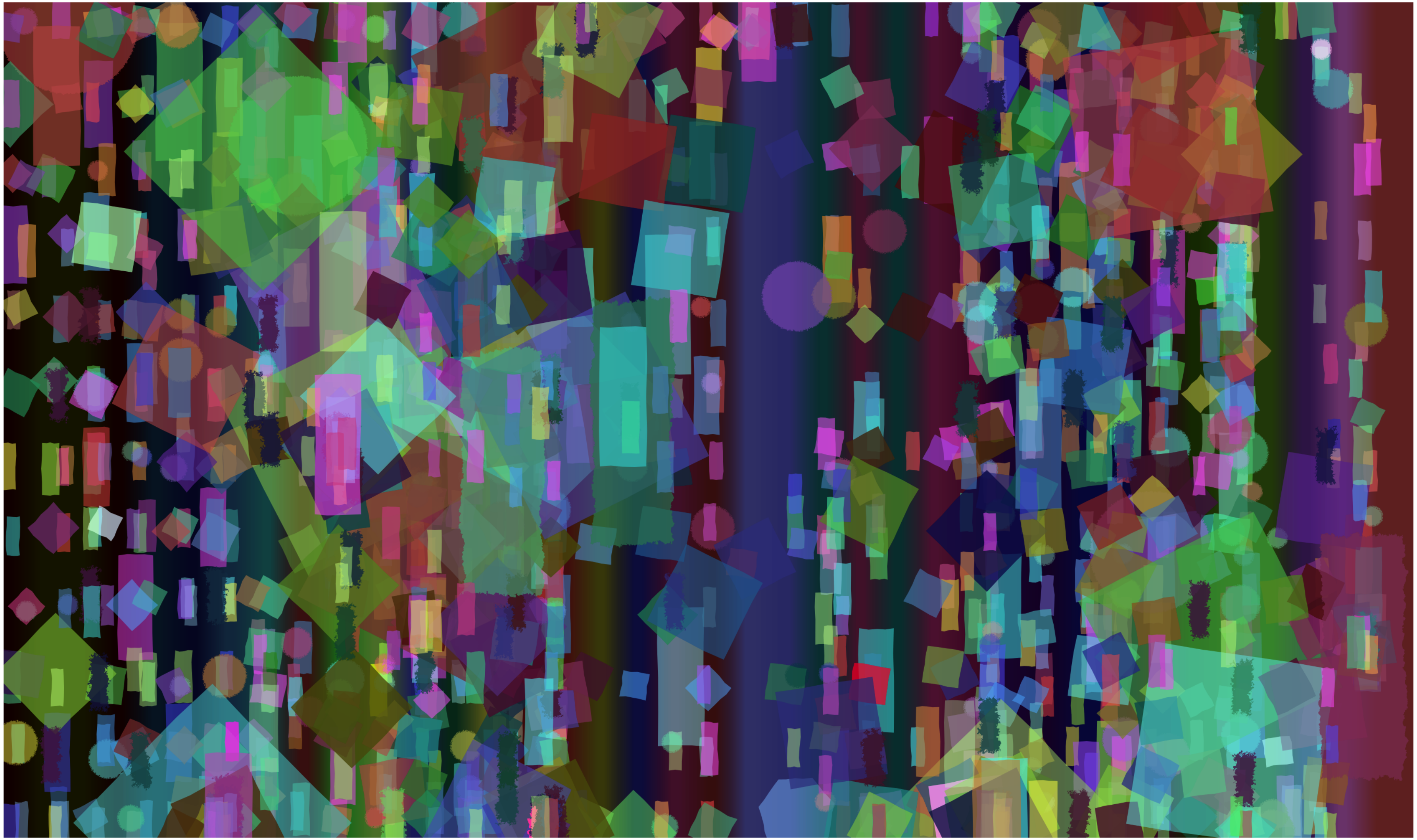


Figure A.12: Stravinsky - *The Firebird*: Genetic Image.