From Music to Image
a Computational Creativity Approach

Luís Ricardo dos Santos Aleixo
luis.aleixo@tecnico.ulisboa.pt

Instituto Superior Técnico, Lisboa, Portugal
November 2020

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

1 Introduction

Computational Creativity (CC) emerged as a subfield of Artificial Intelligence (AI), exploring the machine's ability to generate human-level creative artifacts, hence it focuses on the development of software that exhibits behavior that can be considered creative by humans.

Having as motivation processing music and image algorithmically, our main contribution is the development of a system that exhibits creative behavior by generating abstract images inspired in musical artifacts. We define abstract images as visual artifacts that do not to reflect or convey anything "concrete" or "real". 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.

Regarding the final results, the system outputs 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. Regarding the third version, called the **Genetic** Image, it results from the Genetic Algorithm that receives the two previous versions to generate the initial population.

We decided to assess four pairs of music and respective images through online surveys. The evaluation had good results since the majority of the participants consider the images to be abstract. Besides, regardless of the version, the participants believe that there is a relationship between the images and the music.

Finally, it is also worth mentioning that all generated images were generally liked. Having these interesting results, we believe that our system achieved its goals.

This paper is organized as follows. In Section 2, we provide background knowledge related to the state of the art. In Section 3, we present our background research on the musical domain, and in Section 4, we explore the visual domain by presenting an analysis of different elements of art. In Section 5.1, we explain our approach and in Section 5.2 we describe how we implemented it. In Section 6, we describe the taken approach to evaluate our system, presenting its results. Finally, in Section 7, we present the conclusions of our work, as well as possible future work.

2 Related Work

Diverse research work has been pursued in the Computational Creativity field. Since our goal is to generate abstract images inspired in music, we focus our analysis on inspirational systems in the musical and visual domains.

2.1 Inspirational Systems

The Visual Information Vases (VIV) is an AI-based generative art system proposed by Horn et al. [6] 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.

Teixeira, J. and Pinto, H. [11] 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 visual features, the system interprets and uses them as a starting point to translate into several components of the musical domain. The system outputs three different versions, using one possible approach between the two domains, and Genetic Algorithms to generate music with improved aesthetic value.

Moura, L. [9] 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 [9]. As well as one artist, these robots are stimulated by everything that surrounds them, either the environment, sound, or interactivity with people, leading to the inspiration concept for the drawings.

There are other systems that generate art having as a source of inspiration artifacts from other domains [2, 4]. Since the most used techniques among them are the Learning Systems and Genetic Algorithms, we decided to implement our own in the later, further explained in the following sections.

3 Musical Domain

Music is considered a form of art that usually combines sounds, following arrangements over time - it has a duration and well-defined starting and ending points. In Western Cultures, musical compositions are usually divided into three main parts: **harmony**, **melody** and **rhythm**. 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. Measures are subdivided in individual **notes** - the minimal unit of music - and **chords** - three or more notes played together. It is important to note that this division is relative and that there are elements that rule each of these parts. The following explanation comes from the authors' knowledge acquired throughout this year from other colleagues [1, 11], as well as from music theory sources available online [10].

Parts are generally represented by instruments (or voices) and usually refer to a single strand, or melody or harmony of music, within a larger ensemble musical composition. When analyzing sheet music, each line represents one part of the music, that is, one instrument (or voice).

In Western Music, each staff[1] 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.

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

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 or German terms). However, it can be represented by a numerical value measured in Beats per Minute (BPM).

Throughout the music, each measure has a **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 a measure has. The bottom number is the note value, that is, the type of note to count.

Notes are the minimal unit in musical compositions, and represent sounds that are formed by one mode of vibration of the air. Each note has a **duration** and **pitch**: 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. Besides, the perceptual attribute that enables humans to distinguish among sounds that are playing the same tones, equally loud, is the **timbre**. The difference between two notes is the **interval**, measured in tones. Each note may also be represented by its pitch letter followed by a number, the **octave** in which the note is. It represents how high or low a note should be played.

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.

In Western Music there are twelve ordered pitch-classes[2], designated by the letters of the alphabet, from A to G. These pitch-classes can be represented in a circular diagram, the **Chromatic Circle**. A **scale** is any set of musical notes that can be ascending or descending from the twelve pitch-class, according to its steps. A scale has a tonal center, a key, which is usually the first degree of a scale. In turn, the **Circle of Fifths** 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.

4 Visual Domain

For this work, the visual domain considered is made up of images that do not reflect or convey anything "concrete" or "real" that pertains to elements of the real world.

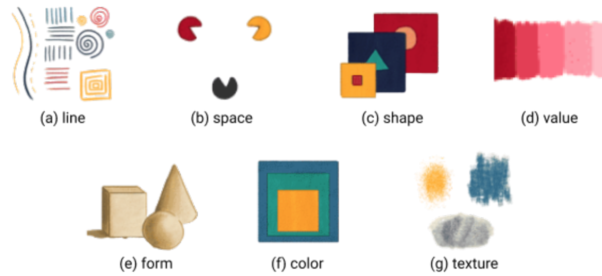


Fig. 1: 7 Elements of Art

According to Esaak, P. [5], there are Seven Elements of Art that serve as "building blocks" to create images on canvas: line, space, shape, value, form, color and texture, as depicted in Figure 1. When generating images, paintings, or drawings to compose a piece of art, these elements may be manipulated and mixed in with design principles.

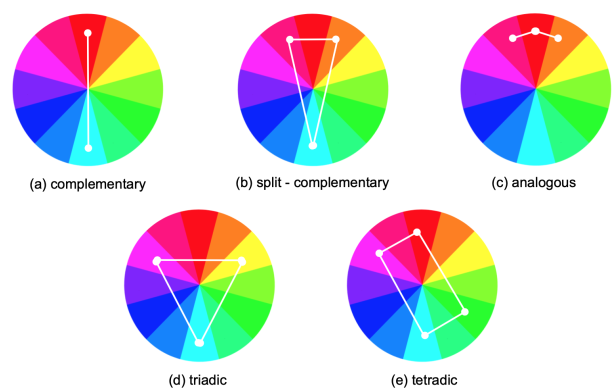


Fig. 2: Color Schemes

Colors are computationally defined by color models, such as the HSV model. In HSV, colors are represented by three measurable attributes: hue, value, and intensity. Hue stands for the color itself, value for the brightness of the color, and intensity for the quality that distinguishes a strong color from a weak one. Colors, as well as music, can be combined to create harmonic results. This harmony is expressed by color schemes, which can be monochromatic, complementary, split complementary, analogous, triadic, and tetradic, as depicted in Figure 2, all of which are based on the Color Wheel [8].

When creating an image, a painting or a drawing, usually a two-dimensional space is generated by a

[2] In musical notation, a pitch-class is the set of all pitches that are a whole number of octaves apart.

shape in two dimensions: height and width. 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. Thus, a **line** is defined by a point moving in space between two points whereby we can visualize stroke movement. **Space** refers to the perspective (distance between and around) and proportion (size) among different elements. **Value** stands for the degree of perceivable color's lightness within an image. **Form** is a three-dimensional element of art that encloses volume - includes height, width and depth. Finally, the **texture** describes the surface quality of the artifact, related to the type of lines used.

These were the basic Visual features we considered for our approach.

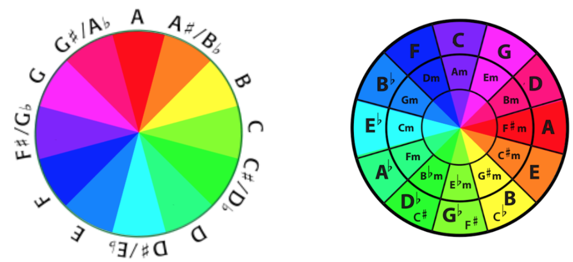
5 Cross-Domain Associations

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.

5.1 Approach

Since we want to be able to perceive all the musical elements in the generated image without the perception of elements' saturation, we decided to analyze music pieces that last between three and five minutes at most. Besides, this way 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, we decided to analyze MIDI files to access the music's basic 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. We also defined the default value for the resulting images as being 1600×2700 pixels, although several dimension sizes for the generated images were tested throughout different phases of the development. 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.

Our approach's foundations lie in the association of the music's **melody** with the image's foreground, and the music's **harmony** with the image's background. In turn, the music's **rhythm** is associated with the size of the image's elements, as well as with the overall luminosity of the image's background. Having our high-level association for the system, we then need to obtain each of the image's elements from the music's structure. We started with the foreground.



(a) Overlapping the Chromatic Circle and the RGB Color Wheel

(b) Overlapping the Circle of Fifths and the RGB Color Wheel

Fig. 3

One of the most basic analogies that can be created between the musical and visual domains is that a note can be translated into a color. Many different associations could be established, even random ones. We decided to overlapped the Chromatic Scale and the Color Wheel - Figure 3a - where each pitch from the 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, that seemed to be a fair enough decision for the mapping we were trying to achieve. The A pitch-class is 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: it is the first color in the visible spectrum, and, in the HSL model, it corresponds to $hue = 0$. To improve the color's quality definition, two other visual properties were added to the hue - saturation and lightness - leading us to have the HSL color model in our mind. 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 saturation 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.

Table 1: One possible shape association for each instrument family

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

At this point, we needed an association to define the **shape** of each visual element. 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. Several different associations were tested. We 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`. One possible association between each instrument's family and shapes is represented in Table 1, where each shape was defined using two elements of the Seven Elements of Art [5] - shape and texture - thus increasing the abstraction level. It is important to note that we made this association based only on 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, in our perception, instruments that do not fit in these rigid shapes. In the end, we decided to consider **two** different approaches for the output of our images: the first with a random shape and texture applied to each instrument found in the sheet music, and the second, where the associations presented in Table 1 were assigned. Time signatures, plus information about *tempo*, are associated with the note's duration to define the element's **size**. These three concepts significantly influence the way we perceive a music piece, and so we decided they should affect the way our image's elements are represented. They were combined to achieve the best responsive size for all the elements that could fit in different screen sizes without the perception of visual elements saturation.

Finally, the **position** of each element in the foreground is directly related to the music's offset - the element's relative location in the music. 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.

According to what is explained in Section 3, one can analyze the music piece as a whole to obtain the most probable tonality or scale used, thus obtaining a sense of its harmony [3]. This way, we performed the analysis of each measure to obtain its most likely tonality key, thus performing an in-depth analysis measure by measure. 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 and the Color Wheel, 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, which is a fair enough decision for the mapping we are trying to achieve. Furthermore, 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 3b. Likewise to the previous case, we decided to consider the HSL color model, thus two more visual properties can be added - saturation and lightness - to the hue to improve the color's quality definition. Since each measure's harmony is composed of chords, we decided to get both the volume and each tonic note [3], to retrieve its octave. The volume is associated with the color's saturation, while the octave defines the color's lightness, thus establishing a possible analogy between chords and colors.

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. Therefore, having the measure's respective color, our background consists of equally sized vertical stripes, whose position is directly mapped from the measure's offset. 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) is 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.

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 5.2.4.

5.2 Implementation

Our system was developed in Python 3.8[3]. To perform the extraction of musical elements, the Music21 [3] library was used. For the generation of the pixels in the image, the Pycairo[4] module was used.

The architecture of our system is divided into several different modules to solve various problems individually, each containing different methods for a specific goal.

5.2.1 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. With this, we perform a top-down analysis for each file, from its parts to its minimal unit, its notes: our system is organized in such a way that we first analyze the Parts, then Measures and finally, measure's Events, in a hierarchical way. When analyzing a **Part**, we get the correspondent Instrument name, as well as its program change[5]. When analyzing a **Measure**, we get its offset. When analyzing a measure's event, if we find an instance of **MetronomeMark**, it means that we are dealing with music's *tempo* for that measure: 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: we get its value. When 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, we get its duration, performing an analysis Note by Note, retrieving the Chord's characteristics.

5.2.2 Cross-Domain Association Module Starting with the part analysis, we get the shape definition. For 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. For the **Associated Image**, the associations represented in Table 1 are applied. Regarding the element's position on the canvas, 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. Each pitch is assigned to the hue definition, among 12 available - Figure 3a. The note's volume, normalized between 0 and 1, is assigned to the color's saturation. The octave is assigned to the lightness and alpha value: after some research and experiments, we consider seven different octaves, from the zeroth to the sixth, and lightness values ranged between 0.1 and 0.7. While the former is a direct translation, the latter applies an inverse operation: the higher the octave, the less solid (more transparent) the element will be: $alpha = 1 - lightness$. Regarding the note's duration, time signature and *tempo*, we assigned them to the element's size, by applying equation 1.

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

Finally, the *tempo* is assigned to the background's overall luminosity and stripe's irregularity by overlap. The music's *tempo* ranges from 60 (slow) to 150 (fast), while the luminosity from 0 (black) to 1 (white), and irregularity from 5 (regular) to 0 (irregular).

Regarding the background, we associate each measure's tonality key to the hue definition - Figure 3b.

[3] https://www.python.org
[4] https://www.cairographics.org/pycairo/
[5] MIDI message that identifies the instrumental sound the device uses when it plays a Note.

For every chord found on each measure, we extract its root note. Then, having all the notes, we calculate an average for the volume to later associate with the color's saturation. In the end, for each chord on each measure, we extract its root note. Similarly, we calculate an average for the predominant octave to later associate with color's lightness and alpha value.

5.2.3 Generator Module Using Pycairo, the system starts by creating a surface with the desired dimensions of the image. Then, it continues by defining the background as a black rectangle with the overall luminosity filter applied. Finally, using a linear gradient, it draws the vertical stripes considering the respective irregularity by overlap, color and position features. To fill out the foreground, since each element has already all the necessary features attached, the system sequentially traverses the array in which the data is stored and, considering each element, draws it using Pycairo drawing methods. In the end, the system outputs their first **two** final versions, the Random and Associated Images, that are saved as separate Portable Network Graphics (PNG) files.

5.2.4 Genetic Algorithm Module To improve the quality of the generated images, we implemented one Genetic Algorithm (GA) mainly applied to the visual domain, although we consider some concepts of music harmony in the generated elements. The algorithm considers an individual an image whose characteristics are the position, color, shape, size of the elements of the foreground, as well as the musical notes that gave rise to them. The initial population of the algorithm consists of 24 individuals. The first two are the Random and Associated Images, and the remaining 22 are obtained from these previous, where the element's horizontal and vertical position in the canvas is shuffled.

For each generation, the fitness of each individual is calculated based on color harmonies between nearest elements, as well as visual perceptions of color, location of elements in the canvas, and shape distributions. For each element we calculate their five nearest neighbors based on the k-nearest neighbors algorithm (k-NN)[6]. Then we score each nearest neighbor according to the harmony between their colors, considering the color schemes presented in Section 4. In turn, we remove score if the color's perception to the human eye is visually close, according to the Delta E distance metric[7], that compares the difference between two colors. This way, we maximize the color diversity of generated elements considering its harmony and visual perception. Then, for each two nearest visual elements, we consider if the musical notes in which they were inspired on are next to each other in the Chromatic Scale. If so, we give a score according to their Euclidean distance, where closest elements are scored higher. Finally, we raise the score if the element's position lies within 10 pixels from its origin position, and if the association between mapped instrument and the element's shape follows the distributions in Table 1. In the end, we remove score if the element's shape is a regular polygon with more than eight sides, since we consider that it reassembles to a circle.

Subsequently, after having a fitness value for each individual, we implemented an **elitism factor**, in which for each new generation, exactly four individuals are passed on to the next generation so that the fitness value never decreases. Then, the selection of each pair of individuals takes into account the **Roulette Wheel Selection**.

In our implementation, on each pair of selected individuals, there is a 90% probability of occurring **Single-Point Crossover** [7], where only the elements' characteristics of one individual are crossed with the elements' characteristics of another individual. These characteristics may be their color, position, shape or size. Then, the **Uniform Mutation** [7] involves changing one element characteristic, and similarly to crossover, several types may occur: Position Mutation, with a 10% chance to occur; Offset Mutation, with a 5% chance to occur; Color Mutation, with a 10% chance to occur; and Shape Mutation, with a 5% chance to occur.

Our algorithm finishes after 300 generations. The final solution is the best population with the fittest individual, that is returned to the Generator Module (Section 5.2.3), thus generating the **Genetic Image**.

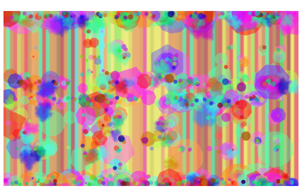
6 Evaluation

Our dataset consists of fourteen music pieces from different eras (from the Baroque to the Modernism) and composers from the Classical tradition. However, it is important to note that our approach is independent of the musical style, and any could have been considered. Although we have executed the system with each of the MIDI files that compose our dataset,

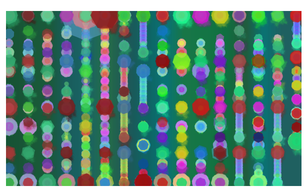
[6] `https://scikit-learn.org/stable/modules/neighbors.html`
[7] `https://python-colormath.readthedocs.io/en/latest/delta_e.html`

for the evaluation process we considered four of them, spread over four online surveys.

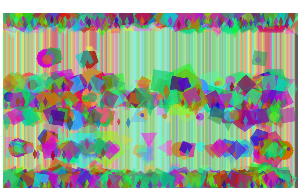
Each survey starts by asking the participants' age, gender, and how frequently they usually go to exhibitions (online, or before the COVID-19 pandemic). Afterward, the survey asks the participant to describe each generated image, using three sentences of their own. Besides, it demands the participants to rank from the most to the least preferred images, and whether participants consider them to be abstract. After that, the survey requests a description of each image version according to predefined adjectives: Exciting; Smooth; Happy; Enjoyable; Surprising; Contemptible; Sad; Aggressive; Disgusting; Boring; Angerful; Fearful. Then, the survey continues to the musical domain, repeating the process that was done with images. After that, for the preferred and least preferred images, the survey asks whether participants generally like them, and if they perceive any relation with the music artifact that served as inspiration for the creation. Finally, it cites the following sentence - "All the images were created having the previous music as inspiration." - and asks if the participants agree with it. These questions were evaluated with a Likert scale from 1 to 5.



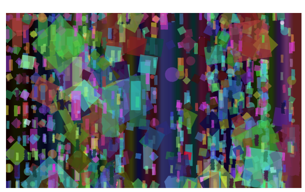
(a) Handel - *Concerto No. 1*



(b) Handel - *The Royal Fireworks Suite*



(c) Mozart - *Symphony No 40*



(d) Stravinsky - *The Firebird*

Fig. 4: Some system's output Images used in the evaluation process

Figure 4 displays the Genetic Images used in the evaluation process. In total, 93 people - 62 female and 31 male - were questioned through all the surveys. Starting by the age group, the main one was $18-29$. Regarding how often participants go to exhibitions, the majority answered: "Once a year".

Table 2: Which version do you prefer?

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 3: Statistics for the most preferred version

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 4: Statistics for the least preferred version

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

The relative frequency for each music and respective preferred image version can be seen in Table 2. From this, we conclude that the most preferred image version, in the majority, is the Genetic one.

The mean, mode, and standard deviation for the results of the most preferred image can be seen in Table 3, while the statistics for the least preferred image can be seen in Table 4. At this point, we related these statistics with the adjectival description participants had on the artifacts.

Regarding the first case, Handel - *Concerto No. 1* and respective preferred version, 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, which can be confirmed by the results in Table 3. The mean and mode value 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 and can be confirmed in Table 4.

The second case is Handel - *The Royal Fireworks Suite*. Here, by looking at the results, participants mainly perceive the image as happy and smooth, while the music as enjoyable and smooth. These results can be confirmed in Table 3, where the mean value for the image rating is 3.7, while the median and mode value is 4. Regarding the relation with the music, since the mean value is 2.93, median value 3, and mode value 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. Once we are assessing the least preferred version, this similarity on the descriptors was not expected. However, by looking at Table 4, for both questions, we observe that participants enjoyed both the images, but perceive a weaker relation between the music and the least preferred image. Anyway, these results are positive, since the chosen descriptors for both artifacts mostly match.

The third case is Mozart - *Symphony No. 40*. Regarding the preferred version, the used descriptors and statistics are similar to the preferred version of the second case. 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 4, where we observe the lower values for the image rating. Nevertheless, we consider that participants perceive the relation between the artifacts.

The fourth and last case is Stravinsky - *The Firebird*. As in the first case, 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. Considering the statistics regarding the relationship between both artifacts - Table 3 - we observe that the mode value 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 results, since the mean value for the preferred image is 3.16, and mode value 5. The least preferred image was perceived equally exciting, happy and enjoyable. However, its statistic values - Table 4 - for the image rating are lower than the preferred version, which was expected. Nonetheless, in comparison with the preferred image, participants consider there is more relation between the image and music.

We cannot fail to point out that, despite the individual analysis, all the images were generally described with adjectives that belong to a positive type of emotions, such as happiness, excitement, or joy. 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. 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 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 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: by observing the statistics, we can see that the majority of the participants believe that there is a relation between the music and image.

Since our main objectives were to generate abstract images that have as inspiration musical files, we consider that the results are desirable given our goals. 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 positive, since the average for the relation between the artifacts is above 3. Finally, regarding the preferred images, the ratings are good - generally, participants seem to like the artifacts, as already seen. Regarding the least preferred images, the results are more neutral, but still good. From these, we can state that participants mostly liked all the images, even the least preferred ones.

7 Conclusions

Inspired in the algorithmic way of processing music and image, the main objective of this work was to develop a system that exhibits creative behavior through inter-domain associations, by generating abstract images inspired from music. 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. Our high-level approach is based on the association of 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.

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 musical composition. The second, called the **Associated** image, was generated assigning a predefined association between instruments and respective shapes. This was entirely based on our personal perception - it is one possible approach for the association process, among others that could have been taken. 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 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.

Regarding evaluation, the majority of the participants considered that the generated images are abstract, and 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 favorable since they go in line with our goals.

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 image filters, as well as to give the user's the possibility to parameterize some variables related to the dimensions of the image.

As future work we consider that 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. Moreover, the instrument's timbre found on MIDI files is very poor, and some essence of the musical composition can be lost in the association process. This way, we propose to use MP3 files instead, so that real versions of each music could be analyzed.

References

1. Francisco Braga and H Sofia Pinto. Sculpture Inspired Musical Composition : One Possible Approach. Master's thesis, Universidade de Lisboa: Instituto Superior Técnico, 2019.
2. Simon Colton. The Painting Fool: Stories from Building an Automated Painter. In *Computers and Creativity*, pages 3–38. 2012.
3. Michael Scott Cuthbert and Christopher Ariza. music21: A toolkit for computer-aided musicology and symbolic music data. 2010.
4. 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.
5. 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.
6. 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.
7. John McCall. Genetic algorithms for modelling and optimisation. *Journal of computational and Applied Mathematics*, 184(1):205–222, 2005.
8. Jill Morton. Basic color theory. `https://www.colormatters.com/color-and-design/basic-color-theory`, Accessed: 2020-08-01.
9. Leonel Moura. Robot Art: An Interview with Leonel Moura. In *Arts*, volume 7, pages 28–28, 2018.
10. Rory Pq. Basic music theory for beginners - the complete guide. `https://www.paulekman.com/universal-emotions/`, Accessed: 2020-08-24.
11. Joana Teixeira and H Sofia Pinto. Cross-Domain Analogy: From Image to Music. In *Proceedings of the 5th International Workshop on Musical Metacreation*, 2017.