Cross-Domain Analogy: From Image to Music

Joana Teixeira, H. Sofia Pinto
INESC-ID, Instituto Superior Técnico
joana.teixeira@tecnico.ulisboa.pt, sofia@inesc-id.pt

Abstract

In this paper, we describe an attempt to create a bridge between the visual and the musical domains. Our system generates a musical artifact given an image as input. To create this bridge (1) we identified a set of features to be extracted from an image, and how these could be interpreted; (2) we took musical theory to understand what features are needed to actually create music; (3) we used the identified visual features and related them to the musical ones. In our implementation, we extract accessible visual features, we interpret them and use them as a starting point. This information is then translated into several features of the musical domain. Two types of output are generated: a raw version, where the visual features are directly translated, and a harmonized version, where some musical conventions are imposed, to create a more aesthetically pleasurable musical artifact. Our current results are very promising since the vast majority of listeners classifies both versions as music.

Introduction and Motivation

Humans are known to be inspired by their surroundings, which may result in the creation of something new. Although a relatively common process in our minds, only recently have researchers attempted to model and implement it in a program. Work in computational creativity focuses both on a theoretical perspective, such as the creation of models and frameworks to describe the creative process, or on a practical perspective, by developing systems that exhibit creative behavior. Both are important, and usually the latter is based on the results of the former. Among many possible sub-areas of computational creativity, cross-domain analogy, or inspiration, is still a mostly under-explored concept (Horn et al. 2015), even though we humans are usually inspired by our surroundings during our creative processes.

Mel Rhodes proposed the "four P's of creativity" – **Process**, **Product**, **Person**, and **Press** (Rhodes 1961). These terms can be used to loosely define creativity as "A process, executed by a person, pressed by his/her environment, by which a product is generated". Therefore, inspiration can be considered the **press** aspect of creativity, since Rhodes defined it as how the environment affects the mind during the creative process.

Taking this into account, our goal was to create a system, which uses visual input (such as an image, a photograph or a painting) as inspiration for generating music. The development of our system required us to better understand each of the domains, individually. Several issues were analyzed: What features can we extract from an image? How can we attribute a specific emotional state to an image? What are the processes of composing a music? What are the different features that a composer can use to create a specific emotional state in a music? Based on the answers to these questions, we created a bridge between the two domains. It should be noted that we do not consider the behavior of our system as purely creative. It uses the creativity that already exists in a visual artifact, and attempts to generate music from it.

Our main contribution to computational creativity is one possible translation, out of many different ones, that uses the image as a starting point to generate a musical artifact. By analyzing features in both domains, we have established possible relations between them. At this phase our main goal was to implement our cross-domain analogy, generate an artifact with it and try to understand if people considered it to be music. Furthermore, since we chose to to apply several rules specific to Rock, our secondary goal was to have our generated artifacts identified as Rock music.

In the next section, we provide some insight on previous work on inspiration and musical generation. Further on, we describe the features we have identified, both in the visual and the musical domains. Then, we explain how we have related these features to each other, creating our cross-domain analogy. In the following section, all the implementation details are explained. Then, we discuss our current results and try to understand if our generated artifacts are, in fact, considered as music by other people, and its genre. Finally, we conclude with some final notes on what we have achieved, some possible applications for our system and future work to further improve it.

Related Work

The main focus of our work is the act of **Inspiration**, which happens frequently in our minds. However, neither psychology (Thrash and Elliot 2003) nor computational creativity (Horn et al. 2015) have dedicated much effort to the study of this topic. Thrash and Elliot (Thrash and Elliot 2003) posit that there are two important objects during an act of

inspiration: the trigger – the object that evokes inspiration – and the target – the object to which the motivation is directed. Thrash and Elliot also wrote that there are three possible sources of inspiration: Supernatural (divine inspiration); Intra-psychic (our own sub-conscious thoughts inspiring our consciousness); and Environmental (nature, other people or works of art).

Inspirational Systems

While psychology attempts to understand and explain what inspiration is and how this process occurs, computational creativity attempts to simulate it with software. The goal is to create computational systems that receive a given input (from any domain – visual, musical, textual, etc) and generate a new artifact that tries to have the same mood, or characteristics (which may or may not be in the same domain).

Johnson and Ventura (Johnson and Ventura 2014) created a system that composes musical motifs, which they describe as "the smallest structural unit possessing thematic identity". To create these motifs, non-musical media is used as inspiration, including non-musical sounds, such as a bird chirping or a running engine, and images. The created motifs may then be used by a human composer to create a full composition, for example. The input is analyzed, and several different candidate motifs are generated. To select one motif as the final output, the authors implemented six different models, which are variations of Markov Models and Neural Networks.

More recently, Horn et al. (Horn et al. 2015) developed a system – Visual Information Vases (VIV) – aiming at capturing and modeling creative inspiration. They define inspiration as the interpretation of concepts from one domain, and their translation into a different domain. VIV receives an image as input, and uses it as inspiration to create 3D-printable vases. The colors of the image are analyzed and used to generate a color palette, with the most predominant colors in the image, to a maximum of eight. The palette is used to calculate four characteristics: hardness, activity, warmth and weight. Together, these define the aesthetic profile of the image. A Genetic Algorithm is used to create the vase. The fitness function attempts to approach the aesthetic profile of the vase (the same four characteristics defined previously) to the one of the image.

Music Generation

Different types of algorithms can be used to generate musical artifacts. Papadopoulos and Wiggins (Wiggins et al. 1999) identified several of the most commonly used algorithms:

- Mathematical Models, used for example in the M.U. Sicus-Aparatus system (Toivanen, Toivonen, and Valitutti 2013);
- Knowledge-Based Systems, used by Oliveira and Cardoso in their Emotion-Driven Music Engine (EDME) system (Oliveira and Cardoso 2010);
- Grammars, used by Steedman (Steedman 1984) in the generation of chord progressions;
- Learning Algorithms, used by Johnson and Ventura (Johnson and Ventura 2014), in the previously mentioned system;
- Genetic Algorithms, which were used by Scirea et al. in the Scientific Music Generator (SMUG) system, in combination with Markov Chains (Scirea et al. 2015).

The first four tend to generate artifacts that follow a certain genre, as they either are trained with music in a specific style (which is the case with Mathematical Models and Learning Algorithms), or they implement and follow rules in the composition process (Knowledge-Based Systems and Grammars). On the other hand, Genetic Algorithms are able to create new and unexpected solutions, although they tend to lack structure, which contrasts with human behavior when composing music. Recently, some researchers have used Multi-Agent Systems to generate musical artifacts, Navarro et al. (Navarro, Corchado, and Demazeau 2014). For inspirational systems, we have seen the use of Genetic Algorithms, Neural Networks and Markov Models in the literature (Horn et al. 2015; Johnson and Ventura 2014).

Identified Features

To create a bridge between the visual and music domains, we had to identify the elements that are readily available or are required from artifacts from each of the domains. The starting point for our system are digitalized images and the output is a midi file. From these elements we try to extract the emotional mood present in the image and map it to the elements that can contribute to the same emotional mood in the music domain. In the following subsections, we summarize what we have learned about both domains.

The Visual Domain

Several different elements can be identified in an image: color, lines, space, texture, shapes, among others. A human can easily look at an image, identify these elements, create his interpretation and try to come up with an understanding of the image as a whole. For a computer, this is still a very challenging task. While it can easily identify the colors and lines, the task of actually recognizing objects, people and animals, understanding their relations among each other and how they are positioned is still a challenge in computer vision (Szeliski 2010). Although extracting the semantic meaning in an image would have contributed greatly to our work, this is not possible, so we have focused on creating a more abstract interpretation of the image by extracting features from the colors present in it. It should be noted that the interpretation of an image and its colors can be a personal matter and may vary according to the person's culture or background, among others. However, we can still try to create a global generalization of color symbolism (Morton 1997), which most of us will recognize.

Color has three properties: hue, value and intensity. The first is the base color, and all the different hues can be found on the color wheel. The second indicates the lightness or darkness of a color. The last refers to the strength and vividness of a color. When combining these properties, thousands

of different colors can be created. In figure 1, the twelve different hues can be seen. Figures 2 and 3 show how the value and intensity can change a color with a specific hue.

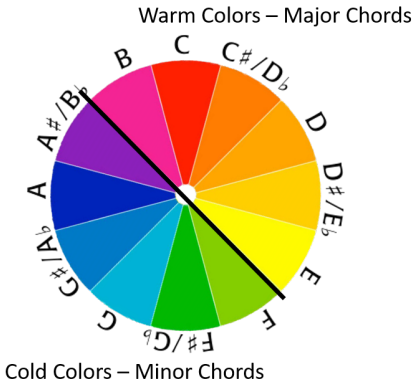

Figure 1: Crossover of the color wheel and the pitch wheel

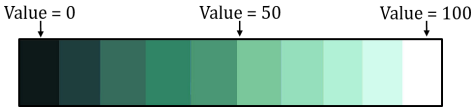

Figure 2: Value scale for the same hue

Figure 3: Intensity scale for the same hue

With this in mind, after obtaining all the pixels present in an image, we can extract different kinds of information:

- Individual pixels give us the specific color of each point in the image.
- A histogram of the image, that shows us most predominant colors of the image.
- How the color varies throughout the image, which allows us to understand if the image mostly maintains the same color, rarely changing, or if it constantly changes colors, creating contrasts.
- A categorization of all the colors into warm and cold, allowing us to understand if the image mostly transmits positive and happy emotions (warm colors), or if it tends to be more negative and sad (cold colors). Figure 1 shows this categorization.
- Groups of pixels which share the same color and are positioned next to each other, creating regions throughout the image.

The Musical Domain

Music is created by bringing together different types of sounds. The final result usually expresses some kind of emotion, and it may or may not be generally considered beautiful[1]. Like described for the visual domain in the previous section, there are also many different elements in music, that can be used to create an emotional mood. The most basic musical element is a **note**, which has four characteristics: duration, pitch, intensity and timbre. By combining several notes together, we can create the main **melody** of a music.

The speed at which a music is supposed to be played is called the tempo, and is usually measured in beats per minute (BPM). Furthermore, a music is always divided into several parts, which are called bars. Bars always have the same duration, and the **time signature** determines how much that is. The rhythm of a music is defined by the tempo and the time signature.

A succession of notes is called a **scale**, and these may be ascending or descending. The name of a scale is always given by its first note, and it may be **(M)ajor** or **(m)inor**. When playing three, or more, notes simultaneously, a **chord** is formed. These can also be major or minor, depending on the notes that are present. By playing several notes together, **harmony** is created. This may be dissonant, if the sound is harsh, or consonant, if it is smooth. When playing in a major scale, or using major chords, the resulting music tends to have a happier mood, whilst minor scales and chords tend to result in more negative moods. Oliveira and Cardoso (Oliveira and Cardoso 2010) identified some possible relations between several musical elements, and the resulting emotional mood. We have summarized some of these, and added some details, on Table 1.

	Happy	Content	Angry	Sad
Scale	Major	Major	Minor	Minor
Volume	Med-High	Med-Low	High	Low
BPM	Med-High	Med-Low	High	Low
Note Density	Med-High	Med-Low	High	Low
Note Duration	Med-Low	Med-High	Low	High

Table 1: Relation between the musical features and the emotional mood.

Finally, a music tends to have a **structure**, with specific instruments and elements. This usually depends on the chosen genre, as some have very rigid structures, whilst others may not even have a defined structure. For example, classical music played by an orchestra uses many different instruments, whilst pop and rock music typically use three or four different instruments. Each genre specifies its own set of rules on how to combine the previously specified elements. It is possible to compose music by ignoring most of these rules, which results in the creation of music that may

[1]The visual domain can also be evaluated aesthetically. However, we only use an image as a starting point. As such, its aesthetic value does not matter to us.

not be considered beautiful by the general public. On the other hand, by following the defined rules, the created music tends to be catchier and more easily liked. For our work, we have decided to generate music that somewhat follows the rules of the rock genre. This simplifies some important decisions, further discussed in the next section. Rock music usually has a drum beat, an electric guitar, an electric bass and sometimes a piano.

Cross-Domain Analogy

Having identified the relevant elements from each domain, they need to be mapped one to the other meaningfully. Since we have chosen to generate Rock music, we have:

- Rock music typically has between 90 to 110 BPM. However, it is possible to go both slower and faster than that interval. As such, we have decided to generate music between 50 and 150 BPM, which allows us to differentiate between extremely varied images and those that have very few color variations.
- A music usually has a main melody and, as such, we need to create a sequence of notes.
- Usually, there are chords played throughout a Rock music. We typically find progressions of three to four different chords being played in a music, in an electric guitar or on a piano.
- Finally, a drum track is imperative in the creation of rock music.

The first value to determine is the number of BPM in the generated music. We considered that an image with more color variations corresponds to a faster music, and vice-versa. As seen in figures 4 and 5, with low color variations there is a feeling of slow movement, whilst higher variations give us a fast-paced movement. As referred previously, a music is divided into several sections with the same duration, which is determined by the time signature. Rock music tends to be composed with a $\frac{4}{4}$ time signature, so we have determined that every musical artifact generated by our system belongs to that time signature.

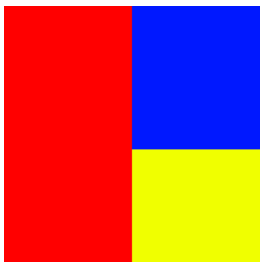

Figure 4: An example of low color variation

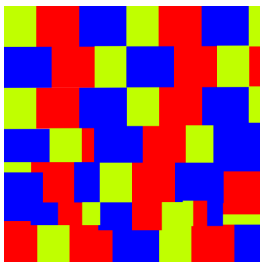

Figure 5: An example of high color variation

Another important mapping is that a color can give us a note. Considering that a color is defined by: (1) its hue – the color itself; (2) its value – its darkness or lightness; and (3), its intensity – how dull or vivid it is –, a direct translation can be made to the pitch (the note itself and its octave) and volume of a note:

1. First, the hue of the color is analyzed. There are twelve different hues in the color wheel, as there are twelve different notes in the chromatic scale. Each one of these hues can be directly translated into a note – for example, a red hue corresponds to a C. The chromatic scale can also be arranged into a circle, so the last note of a scale is also its first. Figure 1 shows how these two circles are overlapped. This translation was chosen since it is the one that makes the most sense to us, given our background. The red hue can be regarded as the first one, just like we consider C to be the first note. A different order would result in a totally different translation.
2. Then, the register of the note should be selected. A higher register corresponds to a higher pitch, and vice-versa. The value of the color gives us the register of the note. So, continuing our previous example, if our red hue has a value of fifty, then our C note is in the fourth octave (generally written as C4).
3. Finally, the intensity of the color gives us the volume of the note. If a color is dull and grayed, then the corresponding note should have a low volume. Likewise, a bright and intense color corresponds to a higher volume.

However, a normal resolution image has far too many pixels to allow us to create a direct pixel-to-note translation, as this would generate very long artifacts. To mitigate this, we have condensed pixels that represent the same color and that are together in the image into ”same colored regions”. These regions are translated into one note, and its duration is directly related to the number of the pixels in the region. This process is explained in more detail in the next section. The analogy of translating a color into a note is also used to choose the chords in the music. For example, if red is one of the most predominant colors, then we have several C chords played throughout the music.

In the visual domain, we have discussed that colors can be categorized into warm and cold colors. Colors that belong to the former tend to feel more vivid and positive, whilst colors from the latter tend to give a more calm and negative feeling. Likewise, a major scale in music transmits positive and happy emotions, while a minor scale is usually more sad and negative. This results in an obvious analogy – if more warm colors are present, then the resulting music mostly uses major scales and chords; if more cold colors are present, more minor scales and chords are chosen. Figure 1 shows which colors are cold, and which are warm, and which chords are major and minor.

Finally, the information presented on table 1 can also be applied to the percussion of a music. As such, we summarized how the visual elements influence the emotional mood of an image on table 2. To create a drum beat for our artifact, we need to evaluate these elements of the image to understand its emotional mood. Then, the musical elements need to be selected to create the closest emotional mood possible. For example, if an image shows a happy mood – high warmth, a relatively high number of color variations, a high color intensity and value –, then the resulting drum beat should have a medium-high volume, BPM and note density, and a medium-low note duration.

	Happy	Content	Angry	Sad
Warmth	High	Med-High	Low	Low
Avg. Region Size	Med-Low	Med-High	Low	High
Color Intensity	High	High	Medium	Low
Color Value	Med-High	Med-High	Low	Med-Low

Table 2: Relation between visual features and emotional mood.

All previously described relations are summarized on table 3.

			Main Melody		Scales	Rhythm		Percussion	Harmony
			Notes	Total Duration	Major vs Minor	BPM	Time Signature		Chords
Image		Image Dimensions		X			4 / 4 (Rock)		
	Color	Individual Pixels	X						
		Variations				X		X	
		Histogram						X	X
		Warm and Cold Colors			X			X	X
		Same Color Regions	X						

Table 3: Cross-Domain Analogy

Implementation

First, it is important to consider one major difference between the visual and the musical domains: whilst the latter has a specific order, a start and a finish, the former does not – an image tends to be seen as whole. However, a computer reads and processes the pixels in a given order. Whichever that order is it, influences the final results. As such, we give the user the possibility to choose the starting point and the order in which they are read. Currently, there are eight different choices, which include starting from the four different corners, and reading left to right, right to left, top to bottom or bottom to top. Figure 6 shows two examples of different starting points and reading orders.

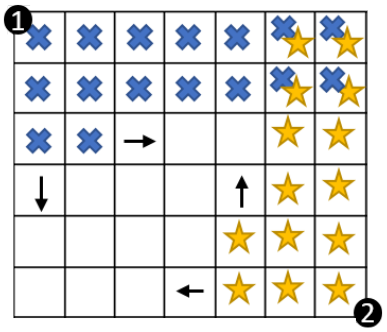

Figure 6: Image traversal examples.

A second concern to be taken into account is how much we should change the generated artifact to comply with a possible aesthetics function. On one hand, we can directly apply all the previously discussed analogies, making a direct translation of the image into a music. However, the resulting artifact may not be considered as music by people, since, in general, music harmony rules are not satisfied by the direct pixel translation. On the other hand, we can obtain an abstract profile of the image, such as the average warmth, average intensity, among others, use these to calculate the musical features, and restrict them to better follow harmonization rules. For example, if the most predominant color is red, then we should mostly use C major chords throughout the music. However, instead of using the second and third most predominant colors to choose the other two chords, we could choose F and G major, as C–F–G is a common rock progression, and it is a known fact that these three chords result in a pleasant harmony. A question that arises, however, is if we can still consider this generated music as the music of the image, that is, if after applying some restrictions, the resulting music is still inspired by the features present in the image and can still be related to it. As such, we decided to create two implementations:

- A ”raw” version, where a pure translation of the visual features into the musical features is attempted.
- A harmonized version, where the visual features are processed, taking into account the whole image, and where harmonization rules are applied in an attempt to ”beautify” the resulting music according to the aesthetic characteristics of the chosen structure.

In this section, we explain how we implemented each one of the two analogies. When relevant, the differences between the raw and the harmonized versions are explained.

The first feature to be calculated is the number of BPM of the music. By using the CIE 2000 color-difference formula (Luo, Cui, and Rigg 2001), it is possible to determine how many times there is a significant color variation in the image. To determine these variations, and then calculate the number of BPM we start by analyzing the image line by line, and each time a significant color difference is found we count a new region. When we reach the end of each line, the average size of same colored regions is calculated, by dividing the number of pixels in a line by the number of regions in that line. This is repeated for every line in the image. By considering all line we calculate the average size of same colored regions for every line in the image. A percentage of the average region size is calculated in relation to the total line size. This is repeated for the columns in the image. When we have the percentages for both lines and columns, an average is calculated between these two values. Finally, this value is translated into the number of BPM. The higher the average size of the same colored regions, then the less variations in the image. This results in a lower BPM value. This translation uses a linear function that maps the percentage value to the number of BPM. The domain of this function is [0, 100], and the range is [50, 150], since we generate music between these BPM values.

After having the number of BPM, the image is divided into several sections, all with the same size, called quadrants. Just like every bar in a $\frac{4}{4}$ music has four beats, each quadrant

in our image corresponds to four beats. The following steps are taken to calculate the dimensions of these quadrants:

1. All the divisors of the height and all the divisors of the width of the image are calculated. This allows us to know into how many parts we can divide the image horizontally and vertically. The result is two sets of numbers. For example, an image sized 15 by 10 pixels has the following divisors, respectively: (1, 3, 5, 15) and (1, 2, 5, 10).
2. All the combinations between these two sets give us the possible dimensions for the quadrants. We are not interested in quadrants that are sized one, nor in quadrants that are the size of the image, therefore, in our previous example, this results in (3, 5) and (2, 5), where all possible dimension combinations are: (3, 2), (3, 5), (5, 2) and (5, 5).
3. To select the best combination, two criteria are followed: The resulting artifact should not be longer than three minutes, nor shorter than one minute, and the width and height of the quadrants should be as close as possible. The total number of quadrants is calculated by dividing the total number of pixels by the number of pixels per quadrant. The total duration of the musical artifact is calculated by multiplying the number of quadrants by four, as each quadrant corresponds to four beats.

Each of the resulting quadrants will have its own histogram from where we can extract several features:

- Most predominant colors in the quadrant.
- Average intensity of the colors in the quadrant.
- Average value of the colors in the quadrant.
- Average warmth of the colors in the quadrant.
- Same colored regions.

To generate the melody of the artifact, we use the basic analogy from color to note explained in the previous section. In our current implementation, there is no difference between the raw and the harmonized versions. Each same-colored region in a quadrant region generates a note, according to its color, and the duration of the note corresponds to the size of the region. So, for example, assuming we have figure 7 as a quadrant, then the following notes are generated:

- D♯, starting at beat 0, and with a duration of two beats;
- F♯, starting at beat 2, and with a duration of one beat;
- C, starting at beat 3, and with a duration of half a beat;
- G♯, starting at beat 3.5, and with a duration of half a beat.

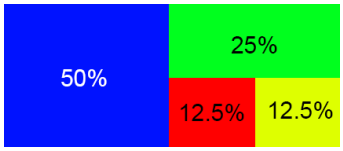

Figure 7: Example of a quadrant divided into same colored regions.

After melody generation, chords are introduced in the music. One chord is played per quadrant, as each quadrant corresponds to four beats in the music. The most predominant color in each quadrant determines which chords are played throughout the artifact, similar to the analogy done in the melody generation. The warmth of that color determines if it is a major or a minor chord. Figure 1 shows the twelve different chords that can be chosen. There are some differences between the raw and harmonized versions:

- In the raw version, chords are selected individually by each quadrant. So, for example, if the most predominant color in the first quadrant is red, a C major chord is played in the first four beats. Then, if the second quadrant has blue as its predominant color, the next four beats have an E minor chord. This is done for every quadrant in the image.
- In the harmonized version, the maximum number of different chords is restricted to four, as a rock music typically uses no more than four different chords. Therefore, we count how many times each color is a predominant color in all the quadrants. For example, in an image with twenty quadrants, we can have a final count of ten for red, five for blue, three for light green and two for purple. The four colors with the highest count are selected, and then are semi-randomly distributed through the duration of the artifact. The chords corresponding to colors with higher counts have a higher probability to be selected, while the others have a lower probability. In our example, red corresponds to a C major and has 50% probability to be chosen, blue is a D minor and has a 25% probability to be chosen, light green is a G with 15% probability and yellow a G♯ with 10% probability of being chosen. This could result in a myriad of different chord orders, which provides a non-deterministic aspect to our program. It should be noted that the final music, in this example, has 20 chords, since the image was divided into 20 quadrants.

Finally, to generate the drum track, we have created a knowledge base of 42 different drum beats used in several popular rock songs. A drum beat is a rhythmic pattern which is repeated throughout a music. Each one of these drum beats is four beats long. For example, figure 8 is a drum beat in our knowledge base, taken from the song "Hotel California", by the Eagles. The emotional value of each of these beats is calculated based on the features presented on table 1, and its selection is different for the raw and harmonized versions:

- If we are generating the raw version, then an emotional value is calculated for each quadrant, based on the features presented on table 2. As such, each quadrant generates a specific drum beat, which has the closest emotional value to that of the quadrant.
- In the harmonized version, we use the general emotional value of the image. It is calculated the same way as in the raw version, however the values are averages for the whole image. First, the closest drum beat is first selected, and is played throughout most of the artifact. However, every sixteen beats (which correspond to four bars in a music), the second closest beat is played, introducing some variety to the drum track.

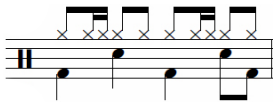

Figure 8: Drum beat from the music "Hotel California", which is played at 72 BPM at a medium-high volume.

Formula 1 shows us how to calculate the emotional mood of a drum beat, where V is the beat volume, NDen is the note density and NDur is the note duration. Formula 2 shows us how to calculate a visual mood, where ARS is the average region size, W is the warmth, I is the color intensity and V is the color value. Both values vary between zero and 100, which makes them directly comparable. Both raw and harmonized versions use the same formulas, however the values used in formula 2 differ as explained above.

$$beat = 0.25 * (BPM + V + NDen + (100 - NDur)) \quad (1)$$

$$visual = 0.25 * ((100 - ARS) + W + I + V) \quad (2)$$

After each of these layers is generated, a MIDI file is created, and each layer is inserted into a different track. Both the chords and the melody are played in piano. Currently, we have generated raw and harmonized versions of several paintings. The generated artifacts can be heard on our website, http://web.tecnico.ulisboa.pt/ist173393/. The system was developed entirely in Python 3.5. To extract all pixels from an image, and to convert these to different color models we used scikit-image[2], a Python module for image processing. To generate the MIDI files, we used the Python module MIDIUtil[3].

Evaluation and Discussion

We gave two short questionnaires[4] to two heterogeneous groups of people. Each survey had a raw and a harmonized version of the same image. Our main goal was to understand if people considered these artifacts as music, and if they could associate them to any genre – more specifically, Rock. For the raw and harmonized versions of the "Broadway Boogie-Woogie" painting we had 58 answers, while having 95 answers to the "Girl Before a Mirror" artifacts.

On average, all artifacts were generally considered as music. The raw version of Broadway Boogie-Woogie had the worst results, with an average of 3.28. Since the medians for both the raw and harmonized versions was four, more than half of the people considered our results as music. In figure 9, all answers to this question are organized into a bar graph.

On the other hand, few people associated the artifacts to Rock. The highest average was the harmonized version of Broadway Boogie-Woogie, with the value 2.03. The medians for the raw and harmonized versions are one and two, which means that less than half people considered these artifacts as Rock. The answers can be seen on figure 10. This

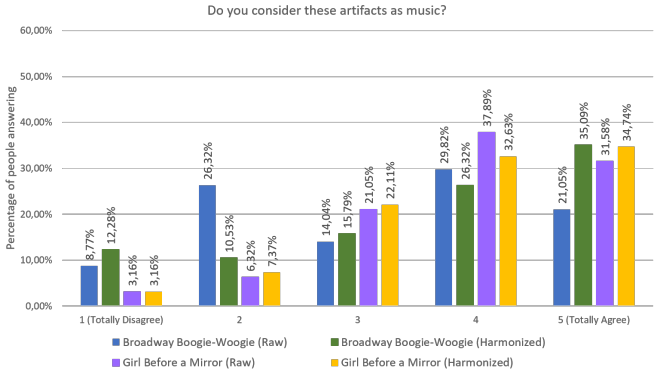

Figure 9: Percentages of answers to the question "Do you consider these artifacts as music?".

may be due to the lack of a guitar track, as most Rock music tends to be identified as such by the use of this specific instrument. However, some people commented that they could identify the artifacts as a baseline to create a Rock music. Other answers identified them as Experimental music, Electronic and even Jazz.

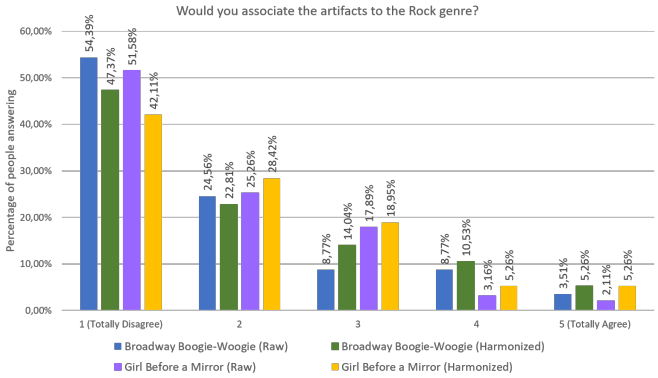

Figure 10: Percentages of answers to the question "Would you associate these artifacts to the Rock genre?".

Since we did not aim at understanding if people related the artifact to the image, we did not explain how it was generated nor its relation to the image. We presented the images nonetheless, so that they could be seen while listening to the artifacts. Even without having any kind of indication that the artifacts were generated from the image, some people commented that the choice of images was interesting and that they could somehow relate them to the music they were listening.

Conclusions and Future Work

In this paper we describe a system that proposes a mapping between the visual and musical domains to generate music from images. From a theoretical perspective, this mapping can be viewed as the use of inspiration for the creation of artifacts. The focus of this paper is on how this mapping can be achieved using easily available features from an image

[2]http://scikit-image.org/

[3]https://pypi.python.org/pypi/MIDIUtil/1.1.1

[4]Questionnaires at: https://goo.gl/forms/aiLE2Sddj6lCjilt2, https://goo.gl/forms/lYkNCD58sZcub4TO2.

and the more consensual, simple and basic features used to compose a music.

From an input point of view, there are no particular constraints imposed on the image to be used as an inspiration. In the case of music, there is a plenitude of music genres. In order to provide a recognizable resulting artifact we chose one particular popular, and relatively simple genre from a composition point of view: Rock. We have used, as examples, images from European Post-Impressionist (Picasso) and Modernist (Mondrian) painters. As this phase we were not dealing with the aesthetic evaluation of the resulting artifact, only with the possibility of being able to generate something that could be recognized as an artifact in the domain – a music. Although these are two very different domains, it was possible to analyze the features of the visual domain, to extract an emotional mood and then to try to replicate a corresponding mood by using the features of the musical domain.

We administered a basic questionnaire to a varied and heterogeneous audience that did not know how the music had been generated. All artifacts were classified as music by the vast majority of listeners. However, the genre was not recognized. We believe that the genre classifications may have suffered from the use of a non mainstream instrument in Rock: we used the piano for both chords and melody. We got very positive feedback from the audience, some even suggesting that a human composer could use the generated musics as inspiration to create more elaborate music from it, which emphasized one of our initial beliefs and goals.

Since the submission of this paper, we have added the electric bass as a new layer to the musical artifact. Furthermore, a third type of generation has been implemented, which results from the execution of a Genetic Algorithm to the two versions described in this paper. A more detailed evaluation was also conducted, where we asked our evaluators if they could relate the generated artifacts to their respective images. We have continued to receive positive results and feedback regarding our system. Even so, more work can still be done, such as the addition of an electric guitar, which has been challenging for us to synthesize with MIDI. Finally, more features in the visual domain can be studied and extracted so that new layers and complexities can be added in the generated musical artifact.

The creation of music inspired by images can be applied, for example, on games that use Procedural Content Generation (PCG). A game could benefit from the automatic generation of music inspired by an image, as each level could have an appropriate soundtrack according to the scenery. This would help the gamer become more immersed in the world, experiencing music that is produced in accordance to what is happening on his/her screen.

Another interesting application of our system would be to accompany a visit to a museum with music inspired by the painting the visitor is looking at. As Compton and Mateas's *Casual Creators* (Compton and Mateas 2015) suggest, not all creative artifacts need to be results of purposeful creative processes, but can emerge from casual, happy, circumstantial, inspiring processes.

Acknowledgments

We would like to thank Miguel Rocha, for providing us with invaluable resources from psychology about creativity and its process, and tips about musical rhythms. We would also like to thank António Gonçalves and Luís Salgueiro, both musical experts from the Calouste Gulbenkian Foundation, for their precious help in the process of composition.

This work was supported by national funds through Fundação para a Ciência e a Tecnologia (FCT) with reference UID/CEC/50021/2013.

References

Compton, K., and Mateas, M. 2015. Casual Creators. In *Proc. of the 6th Intern. Conf. on Computational Creativity (ICCC 2015)*, 228–235.

Horn, B.; Smith, G.; Masri, R.; and Stone, J. 2015. Visual Information Vases: Towards a Framework for Transmedia Creative Inspiration. In *Proc. of the 6th Intern. Conf. on Computational Creativity (ICCC 2015)*, 182–188.

Johnson, D., and Ventura, D. 2014. Musical Motif Discovery in Non-Musical Media. In *Proc. of the 5th Intern. Conf. on Computational Creativity (ICCC 2014)*, 91–99.

Luo, M. R.; Cui, G.; and Rigg, B. 2001. The development of the cie 2000 colour-difference formula: Ciede2000. *Color Research & Application* 26(5):340–350.

Morton, J. 1997. *A Guide To Color Symbolism*. Colorcom.

Navarro, M.; Corchado, J. M.; and Demazeau, Y. 2014. A Musical Composition Application Based on a Multiagent System to Assist Novel Composers. In *Proc. of the 5th Intern. Conf. on Computational Creativity (ICCC 2014)*, 108–111.

Oliveira, A. P., and Cardoso, A. 2010. A Musical System for Emotional Expression. *Knowledge-Based Systems* 23(8):901–913.

Rhodes, M. 1961. An Analysis of Creativity. *The Phi Delta Kappan* 42(7):305.

Scirea, M.; Barros, G. A. B.; Shaker, N.; and Togelius, J. 2015. SMUG: Scientific Music Generator. In *Proc. of the 6th Intern. Conf. on Computational Creativity (ICCC 2015)*, 204–211.

Steedman, M. 1984. A generative grammar for jazz chord sequences. *Music Perception* 2:52–77.

Szeliski, R. 2010. *Computer Vision: Algorithms and Applications*. Springer.

Thrash, T. M., and Elliot, A. J. 2003. Inspiration as a Psychological Construct. *Journal of Personality and Social Psychology* 4(84):871–889.

Toivanen, J.; Toivonen, H.; and Valitutti, A. 2013. Automatical Composition of Lyrical Songs. In *Proc. of the 4th Intern. Conf. on Computational Creativity (ICCC 2013)*, 87–91.

Wiggins, G. A.; Papadopoulos, G.; Phon-Amnuaisuk, S.; and Tuson, A. 1999. Evolutionary Methods for Musical Composition. In *International Journal of Computing Anticipatory Systems*.