The music of paintings: a rhythmic perspective

André Gonçalves, MSc.
GAIPS, Inesc-Id, Instituto Superior Técnico, Universidade Técnica de Lisboa, Lisboa, Portugal
e-mail: andre.pintado@ist.utl.pt

Eduardo Lopes, PhD.
Departamento de Música, Universidade de Évora, Évora, Portugal
http://home.uevora.pt/~el
e-mail: el@uevora.pt

Ana Paiva, PhD.
GAIPS, Inesc-Id, Instituto Superior Técnico, Universidade Técnica de Lisboa, Lisboa, Portugal
e-mail: ana.paiva@inesc-id.pt

Abstract

Olivier Messiaen once mentioned: "I am affected by a kind of synopsia which allows me, when I hear music, and equally when I read it, to see inwardly, in the mind's eye, colours which move with the music". It is now known that a number of artists have a condition called synesthesia. Synesthesia is a neurological phenomenon where the stimulus of a sense induces the stimulus of another sense. For instance, there are people who see colours when looking at some specific numbers, or when hearing some specific music. Since it enables the association of typically non-associated concepts, it is believed that this phenomenon is the key to creativity.

It is easily perceivable that some rhythms convey a more relaxed experience than others. This also applies to shapes. We can call this the instability/stability of a rhythm or shape. Using a mathematical model created by us, we quantify the instability of a shape present in a painting and then, using the theory of Pulse Salience, generate a rhythm with a similar instability value.

We present a computational system which is able to play the music/rhythms of paintings. The system is organized in three layers, and there are two key modules: an image analyzer tool and a rhythm generator tool. The former extracts the image

characteristics and assigns an instability/stability value to a shape, while the latter generates and plays rhythms according to the instability/stability values.

The aim of the system is to provide a multi-sensorial experience similar to synesthesia, using the cognitive and perceptual influence of rhythm in humans. We believe that if people can be exposed to a similar experience, this will improve their overall perception of a work of art as well as enhance their own creativity.

1. Introduction

"If there is no shape, there is no flavor" [1]. This quote could be thought as a comedy quote, but it is not. The man who tasted shapes is a famous example of a person who experienced a condition called synesthesia, a condition in which people mix senses. His mind works like a house with no walls, where the senses do not have their 'privacy' and finish melting with each other. In his case, he mixes taste and touch, but there are many more variants of this condition, some more common than others [2,3].

Transposing this inter-sensorial experiences to art, it is known that artists are often appealed to dig new ground into other areas. Painters like Paul Klee or Wassily Kandinsky often created paintings based on musical rhythms, harmonies or melodies [4,5]. It is also known that musicians often compose or improvise inspired by other media. Whether it's painting, architecture [6], poetry or dance, these composers find a mapping between both fields, so they can cause an effect with music similar to the effect caused by the other media. The concepts mapped can vary from abstract ones (emotions, memories, etc.) through more concrete ones (textures, form/geometry, movement, etc.). Film score and cartoon music composers are experts in translating the real ambiance and feelings of the characters into the music. More generally, there have been art movements, like Fluxus in the 60's, which were entirely dedicated to projects blending different artistic media and disciplines, also called intermedia [7].

So, how can we provide a synesthetic experience to people? More specifically, how can we make the computer generate music based on drawings or images so that the music generated transmits the same feelings as the drawing?

We named our system Sense^2, which reads 'sense square' and has itself multiple interpretations: the multiplication of senses; the reference to two senses; the reference to the 'square' as a geometry figure; and 'sense the square' in a reference to the imperative mood of the verb 'to sense', as if we were strongly suggesting the user to feel the geometry.

Our fields of action are paintings or drawings, and music. We must then map this multi-disciplinarity by finding a common feature in both fields. First, we focus ourselves on the shapes present in the images and on musical rhythms. Second, the bridge between shapes and rhythms is made with a feature we call instability. Instability is the power of something to be or not to be smooth to the perception of a person. In short, an unstable shape is not smooth and has sharp edges, whereas an unstable rhythm has an unexpected variation/surprise on the perception of the people who are listening to it, and depending on its value can even cause confusion.

This document is organized in four chapters. In the first chapter we describe the background concepts introduced previously: music, rhythm, pulse salience and computer music. In the second chapter we present the details behind the implementation of Sense^2. In here, we explain all the details about the mathematical formulas that constitute both models of instability, and also some system programming details. In the third chapter we present the evaluation of our system. Finally, in the fourth chapter we conclude on the results of all our work, and on the results of the evaluation. We finish by proposing future work that could not only improve Sense^2 but also make use of the models in other scenarios.

2. Background

2.1 Music and Rhythm

Music has plenty of characteristics and from the point of view of automatic music generation, one can think of lower level characteristics that would be proper to map: harmony, melody, timbre, etc. Or higher level ones: multiple instruments, patterns, musical styles, etc. For cultural reasons well beyond the scope of this document, in western music theory, pitch has been preferred over rhythm, and the proof is that in today's music education, pitch is more widely spread, and quite more valorized than rhythm. This happens because of the rhythm naturalness in human perception, with humans tending to ignore or underestimate what is natural to them.

In musical terms, rhythm (from the Greek 'rhythmos' meaning "flow" or "movement") is the arrangement of sounds in time. By arrangement we mean the organization of musical notes or events in an interval of time, which can include the duration of notes (and silences), the accentuation of notes, meter and tempo. If one takes rhythm out of a music piece, one is not able to recognize it anymore. But if one takes the pitch out of a piece, one still gets something musically rich (and perhaps sometimes similar to popular dance music). These crucial perceptual qualities cause rhythm to be the most important music characteristic and yet it has not been fully recognized by Musicology and music theory, at least in the shape of rhythm treatises.

2.2 Pulse Salience

The Just in Time theory introduces Pulse Salience. Pulse salience is a rhythm characteristic that characterizes the salience of a given rhythmic event [8,9]. A salient pulse is like the black sheep that stands out in a flock, except that non-experienced listeners internalize it without noticing it. It can be measured through three factors: the duration of a pulse (also called agogic accentuation), the accentuation effect due to the action of rhythm cells or rhythm motifs, and the accentuation resulting from the placing of a pulse over a strong metrical point which is perceived as stable.

2.2 Computer Music

When computers were invented in the 40's, musicians and researchers were interested in using them to assist humans with music tasks such as performing or composing [10,11]. The Computer Music area can be divided into four categories: synthesis, composition, performance and analysis. We focus ourselves in composition.

Computer music composition is all about algorithms. There are literally hundreds of methods and hybrids of methods, created since the beginning of the computer music era, and some of them were even created longtime before computers were born [12,13]. Classification of computer music composition methods often divide these in the following groups: Mathematical Models, Knowledge Based Systems, Grammars, Evolutionary Methods, Learning Systems, Distributed Systems, ‘Sonification’ and Hybrid Systems. This categorization is quite empirical as, for example, we can consider almost all of them to be mathematical models or knowledge based systems. Each one of these groups differentiates from the others not only by their computational features but also by their different use. For a complete and detailed reference on these methods please refer to [10,14,15].

3. Sense^2

Sense^2 is a system which generates music following paintings. For this reason, one must create a direct or indirect mapping between paintings characteristics and music characteristics. Rather than to map concrete painting characteristics into concrete music ones, which could lead us to non-creative results, we have chosen to take a step towards human consciousness and decided to use human feelings as the bridge between both media. We use the fact that some paintings or some music convey a more relaxed experience than others. We focus ourselves on shapes and rhythms, and we call this the instability/stability of a shape or rhythm.

3.1 Measuring Instability

3.1.1 Instability of shapes

The calculation of the instability of a shape depends on whether the shape is a line, a circle or a polygon.

We define the line instability as its inclination, or the angle formed by it and an imaginary horizontal line, where a vertical line (90º) has 0% instability and a horizontal line (0º) has also 0%. The 100% instability is reached when the line forms a 45º angle with the imaginary horizontal line (Figure 1).

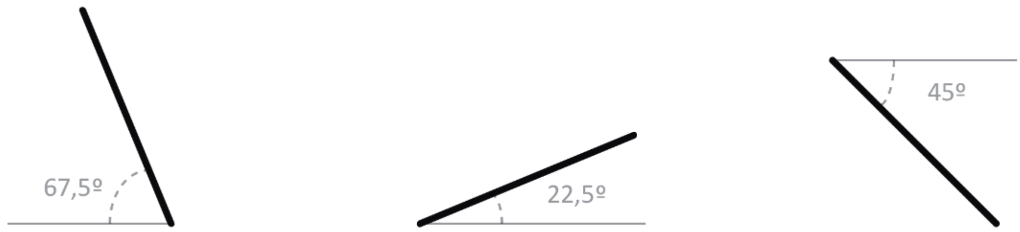

Figure 1a: 67,5º = 50% unstable

Figure 1b: 22,5º = 50% unstable

Figure 1c: 45º = 100% unstable

Figure 1: Example of line stabilities

These angles are calculated using the mathematical operation dot product (where a and b are vectors):

$$\theta = \arccos\left(\frac{a \cdot b}{|a||b|}\right) \tag{1.1}$$

Which, applied to the line points p2, p3 and the imaginary p1:

$$\theta = \arccos\left(\frac{(x_1 - x_2)\times(x_3 - x_2) + (y_1 - y_2)\times(y_3 - y_2)}{\sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2} \times \sqrt{(x_3 - x_2)^2 + (y_3 - y_2)^2}}\right)$$

Then, to obtain the percentage (all the angles are previously normalized from 0º to 90º):

$$instability_l = \begin{cases} \theta \times \dfrac{100}{45} & \text{if } \theta \le 45^{\circ} \\ (90 - \theta) \times \dfrac{100}{45} & \text{if } \theta > 45^{\circ} \end{cases} \tag{1.2}$$

The circle shape is always considered to be a stable shape. For this reason its instability is always 0%.

$$instability_c = 0\% \tag{2}$$

The calculation of the stability of a polygon is the most complex of the three. One can easily see that the smoothness of a polygon depends on angles and on the presence of those angles in the shape. More precisely, here we say that it depends on the number of acute and obtuse angles and on the size of the rays that create those angles. We should add that all the angles are normalized between 0º and 180º, and 90º is considered to be an obtuse angle. Let's look at the method in detail.

Let $\hat{a}_{a1}\ldots\hat{a}_{an}$ be the acute angles of a polygon, and $l_{a1}\ldots l_{an}$ the length of the two rays that form that angle. In the same way, let $\hat{a}_{o1}\ldots\hat{a}_{on}$ be the obtuse angles of a polygon, and $l_{o1}\ldots l_{on}$ the length of the two rays that form that angle (Figure 2).

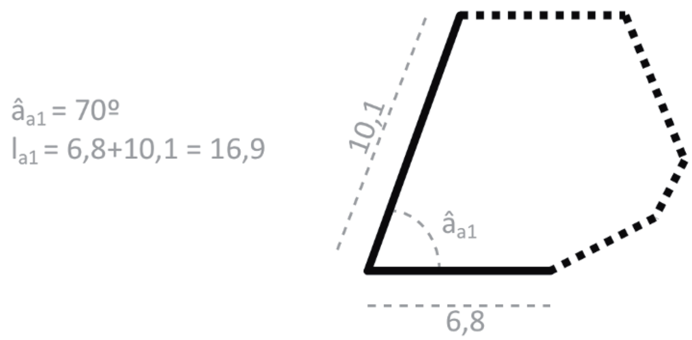

Figure 2: Example of acute angle in a polygon and the calculation of the length of its two rays

n_a and n_o are respectively the number of acute and obtuse angles of a shape. We define the "Acute Perimeter" of a shape as the sum of the lengths of the two rays that form an acute angle multiplied by that angle. The multiplication is used as a way to weigh the angle. An angle formed by two small rays must have a smaller value than the same angle formed by two large rays. All angles are previously normalized from 0º to 180º. Thus, we have:

$$AcutePerimeter = \sum_{k=0}^{n_a} l_{ak} \times (90 - \hat{a}_{ak}) \quad (3.1)$$

$$ObtusePerimeter = \sum_{k=0}^{n_o} l_{ok} \times (180 - \hat{a}_{ok}) \quad (3.2)$$

We then defined the Acute Value and Obtuse Value as:

$$AcuteValue = \frac{AcutePerimeter}{n_a} \quad (3.3)$$

$$ObtuseValue = \frac{ObtusePerimeter}{n_o} \tag{3.4}$$

This operation is a way to weigh the perimeter value which can be quite big. If there are no acute angles the “Acute Value” is obviously 0, and the same occurs with the “Obtuse Value”.

Finally, Instability or “Acute Percentage” is a percentage of the “Acute Value” in the total value composed by Acute and Obtuse Value:

$$instability_p = AcuteValue \times \left(\frac{100}{AcuteValue + ObtuseValue} \right) \tag{3.5}$$

To exemplify our polygon instability model, we present two examples of instability quantification for two different polygons. Let us calculate the Acute and Obtuse Values for the first polygon example (Figure 3):

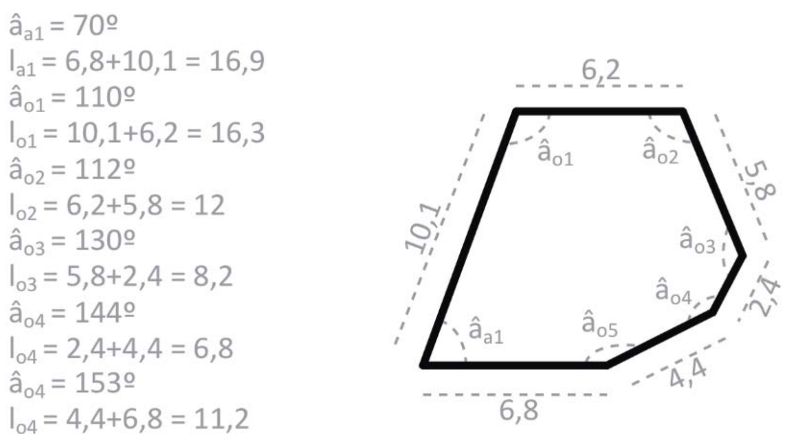

Figure 3: A detailed view of a quite stable shape

$$AcuteValue = \frac{20 \times 19,9}{1} = 338$$

$$ObtuseValue = \frac{(80 \times 16,3) + (78 \times 12) + (50 \times 8,2) + (36 \times 6,8) + (27 \times 11,2)}{5} = \frac{3197,2}{5} = 639,44$$

Then, using formula 3.5, the instability percentage:

$$instability_p = 338 \times \left(\frac{100}{338 + 639,44} \right) = 34,58\,\%$$

Let us do the same process for a more complex shape (Figure 4):

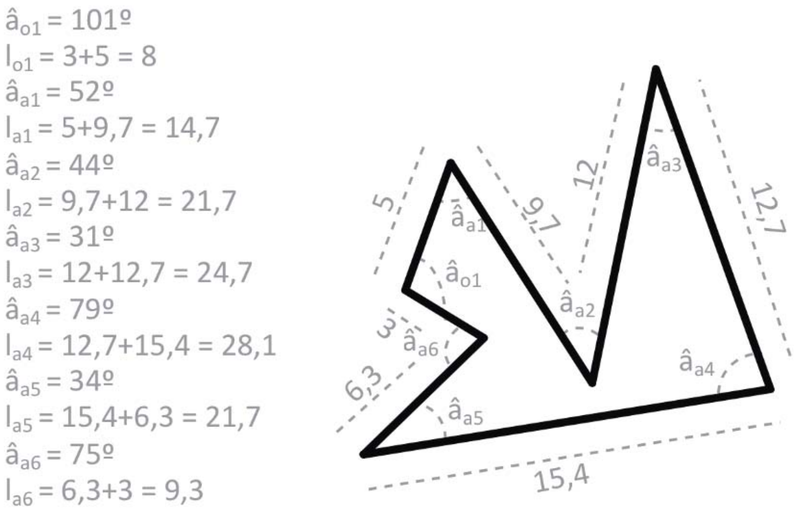

Figure 4: A detailed view of an unstable shape

$$AcuteValue = \frac{(38 \times 14,7) + (46 \times 21,7) + (59 \times 24,7) + (11 \times 28,1) + (56 \times 21,7) + (15 \times 9,3)}{6} = \frac{4677,9}{6} = 779,65$$

$$ObtuseValue = \frac{79 \times 8}{1} = 632$$

Then, using formula 3.5, the instability percentage:

$$instability_p = 779,65 \times \left(\frac{100}{779,65 + 632} \right) = 55,22\,\%$$

One can see that a polygon with no acute angles has 0% instability, and a polygon with no obtuse angles has 100% instability. This is quite arguable, as an equilateral triangle, for example, can be seen as being a stable and unstable shape. This model sees it as being all unstable. Yet, The main goal when making this model was to mathematically reflect the smoothness of the shape, and the two previous examples show that the model is quite accurate.

3.1.2 Instability of rhythms

Our instability quantification method is based on Pulse Salience, a rhythm characteristic which associates values to notes following their salience. Let us first introduce the Pulse Salience method, which is essential to the comprehension of our theory. Then, we will detail our rhythm Instability Quantification method, which is used in the Rhythm Generator and Rhythm Chooser modules.

Pulse Salience

As said before, pulse salience is a rhythm characteristic which characterizes the salience of a given rhythmic event. In Ricardo Cruz's thesis [16,17], the author identified a way to quantify the salience of every note in set of measures, or a score. Like in Lopes' theory, the salience value of a pulse depends on three factors:

- The pulse position in the measure, also called "metric position";
- The pulse duration, also called "agogic accentuation";
- The rhythmic events that precede the pulse which can accentuate the pulse, also called "rhythm cell accentuation".

Recalling the mathematical functions from (Cruz 2008):

Metric Position Value:

$$O(\omega) = BU - M(\omega - 1)$$

Agogic Accentuation Value:

$$P(\eta) = \eta$$

Rhythm Cell Accentuation Value:

$$Q(\eta) = \sum_{(\zeta,\sigma)\in C} \left(\left\lceil R\left(\frac{\eta}{\zeta}\right) M\sigma \right\rceil \right)$$

From which the author derived the Pulse salience general formula:

$$S(\eta,\omega) = BU - M(\omega - 1) + \eta + \sum_{(\zeta,\sigma)\in C} \left(\left\lceil R\left(\frac{\eta}{\zeta}\right) M\sigma \right\rceil \right) \qquad (4)$$

Where:

η is the pulse's type

ω is the pulse's metric position

B the number of beats in each measure

U the value filling a beat

M the value of the shortest pulse present in the rhythmic segment.

C a set of tuples (ζ , σ) representing the preceding rhythm cell - that can be in fact a group of equal homogeneous cells. For each tuple, ζ is the value of a pulse type and σ the number of pulses of that type in the rhythm cell.

R a function given by:

$$R(x) = \begin{cases} x & \text{if } x > 1 \\ 0 & \text{if } x \leq 1 \end{cases}$$

Figure 5 represents all the possible metric positions of a pulse in a measure, and table 1 represents the different pulse types used in our system, and their possible values.

Figure 5: The sixteenth metric positions in a quaternary meter following 'Just in Time'

Pulse	Notation	Value
Quarter Note	♩	1000
Dotted Eighth Note	♪.	750
Eighth Note	♪	500

Sixteenth Note	𝅘𝅥𝅯	250
Sixteenth Silence	𝄿	0

Table 1: Pulse types and their possible values

Here are the pulse salience values for the different pulses for the rhythm in figure 6. For more detailed examples, please refer to (Cruz 2008).

Figure 6: Pulse salience values for a random rhythm

$$S(1000,1) = 4 \times 1000 - 250 \times (1-1) + 1000 + 0 = 5000$$

$$S(500,3) = 4 \times 1000 - 250 \times (3-1) + 500 + 0 = 4000$$

$$S(0,7) = 0$$

$$S(0,15) = 0$$

$$S(250,2) = 4 \times 1000 - 250 \times (2-1) + 250 + 0 = 4000$$

$$S(250,10) = 4 \times 1000 - 250 \times (10-1) + 250 + 0 = 2000$$

$$S(250,6) = 4 \times 1000 - 250 \times (6-1) + 250 + 0 = 3000$$

$$S(250,13) = 4 \times 1000 - 250 \times (13-1) + 250 + 0 = 1250$$

$$S(1000,4) = 4 \times 1000 - 250 \times (4-1) + 1000 + 4000 = 8250$$

As one can see, the quarter note in metric position 4 has a salience value of 8250 especially because it is located straightly after a rhythm cell of four sixteenth notes.

Instability of Rhythms

Our model divides rhythm instability into two mathematical measurements: Pulse Instability and Rhythm Instability. Rhythm Instability may only be calculated if Pulse Instability was calculated previously for all the notes in the measure.

Let us start with Pulse Instability. We say that a note is unstable if it has a high salience value and is positioned in a weak metric position (such as 10 or 16). The Pulse Instability value must mirror this in perfection, so we define it as:

$$instability_{pulse}(\eta,\omega) = S(\eta,\omega) \times \omega$$

This will cause the Pulse Instability value to increase proportionally with metric position. There is however one special case, which is the silence case. Silences haven't got salience value because they do not sound, but the fact is that their presence can cause instability/stability. For this purpose, we have contemplated silences in our formula:

$$instability_{pulse}(0,\omega) = \frac{\sum_{\omega \in s} O(\omega)}{n_s^2} = \frac{\sum_{\omega \in s} BU - M(\omega - 1)}{n_s^2}$$

Where:

s is the group of silences where the given silence is present

n_s is the number of silences that form the group of silences where the given silence is present

The division by n_s^2 works as a weighting to reduce the importance of instability of many silences put next to each other, and to spread the value over all the silences that belong to the group.

The general formula for Pulse Instability is then given by:

$$instability_{pulse}(\eta,\omega) = \begin{cases} \dfrac{\sum_{\omega \in s} BU - M(\omega - 1)}{n_s^2} & \text{if } \eta = 0 \\ S(\eta,\omega) \times \omega & \text{if } \eta > 0 \end{cases} \qquad (5)$$

The Pulse Instability formulas are able to associate an instability value to each rhythm pulse, but there is still a problem concerning the quantification of the overall rhythm instability. This is why we created Rhythm Instability. We cannot just sum the Pulse Instability for each pulse in a rhythm, because this would result in a higher Rhythm Instability value for rhythms with many pulses, and a lower one for rhythms with a few pulses. The solution was to weight the overall Pulse Instability value by the number of pulses in the rhythm:

$$instability_{rhythm} = \frac{\sum_{\omega=1}^{16} instability_{pulse}(\eta,\omega)}{n_{pulses}} \quad (6)$$

As the instability of a shape is defined as a percentage, the rhythm instability must also be defined as such, for coherence reasons. We did so, by dividing the Rhythm Instability values in 10 intervals so that this conversion from an instability value to a percentage could be done. For the definition of the interval endpoints we produced some studies where we generated large quantities of rhythms and noticed that the Rhythm Instability values oscillated in between 6000 and 32600, and most of them between 11320 and 21960. As a result we defined the Instability intervals as such:

0% to 10% : $[6000, 11320[$

10% to 20% : $[11320, 12650[$

20% to 30% : $[12650, 13980[$

30% to 40% : $[13980, 15310[$

40% to 50% : $[15310, 16640[$

50% to 60% : $[16640, 17970[$

60% to 70% : $[17970, 19300[$

70% to 80% : $[19300, 20630[$

$$80\% \text{ to } 90\% : [20630 , 21960 [$$

$$90\% \text{ to } 100\% : [21960 , 32600 [$$

Here is the detailed calculation of the Rhythm Instability for the rhythm present in figure 6:

$$instability_{pulse}(1000,1) = (4 \times 1000 - 250 \times (1-1) + 1000 + 0) \times 1 = 5000$$

$$instability_{pulse}(500,3) = (4000 - 250 \times 2 + 500) \times 3 = 4000 \times 3 = 12000$$

$$instability_{pulse}(0,7) = \frac{(4000 - 250 \times 6) + (4000 - 250 \times 14)}{2^2} = \frac{3000}{4} = 750$$

$$instability_{pulse}(0,15) = instability_{pulse}(0,7) = 750$$

$$instability_{pulse}(250,2) = (4000 - 250 + 250) \times 2 = 8000$$

$$instability_{pulse}(250,10) = (4000 - 250 \times 9 + 250) \times 10 = 20000$$

$$instability_{pulse}(250,6) = (4000 - 250 \times 5 + 250) \times 6 = 18000$$

$$instability_{pulse}(250,14) = (4000 - 250 \times 13 + 250) \times 14 = 14000$$

$$instability_{pulse}(1000,4) = (4000 - 250 \times 3 + 1000 + 4000) \times 4 = 330000$$

$$instability_{rhythm} = \frac{instability_{pulse}(1000,1) + ... + instability_{pulse}(1000,4)}{7} = \frac{111500}{7} = 15928,57$$

And then, following the intervals defined previously, we can see that 15928,57 fits in the instability interval 40% to 50%. So we say that the rhythm in figure 6 has an instability value between 40% and 50%. The reader may have noticed that one cannot say precisely what is the instability percentage of the rhythm, as it is defined by an interval and not by a unique value. Briefly, there is no need to computationally distinguish between a rhythm 41% and 42% unstable, because most human ears would not be capable to do it either. This will be detailed in the next section [18].

3.2 The System

The System Driver is the first module to be executed upon the running of Sense2. He is the responsible for calling the interface system, the picture analyzer module and the synesthetic layer. He guarantees that all these modules are executed in this order. The interface is composed by a “file selection window”, followed by a window where the selected image is loaded. After this file selection, its absolute path is sent to the Picture Analyzer.

The Picture Analyzer was implemented using the openCV library, and it is responsible for extracting lines, circles and polygons from pictures. Lines and circles are extracted using the Hough transform and the polygons are extracted using Canny or Threshold algorithms. All this information is then written to a XML file for later use.

The Synesthetic Layer is responsible for converting image characteristics into instability values. It uses the concepts and model explained in section 3.1.1 (instability of shapes). The instability values calculated by this layer are then sent to the Composition Layer.

The composition process is composed of two main tasks: generation and choice. The Rhythm Instability Function defined previously is not a bijective function, which means that there are numerous rhythms with the same Instability value and that given an Instability Value we cannot get one rhythm. Following this thought we created the Rhythm Generator, a module which generates rhythms.

To simplify the generation and hearing we have introduced some constraints and yet there were still more than 3.000.000 rhythms. A search in this set would reveal itself to be exhaustive and uninteresting, because even a trained listener wouldn’t notice the difference between most of them. We decided then to generate 100 new rhythms every time the program was run. To guarantee that the Rhythm Instability value for these rhythms was spread all over the possible Rhythm Instability values, we divided the Rhythm Instability values in 10 intervals so that each one of the intervals would contain 10 rhythms.

The Rhythm Chooser will choose the right rhythm in this list, following the instability value sent by the Synesthetic Layer. The rhythms are taken from the rhythm list quite directly. For example, if the shape instability percentage is 31%, the rhythms will be randomly taken from the fourth interval (30% to 40%), and if the shape instability percentage is 9%, the rhythms will be taken from the first interval (0% to 10%).

Figures 7, 8 and 9 present some possible rhythms for the shapes on the left.

Figure 7: One possible rhythm for the line on the left

Figure 8: One possible rhythm for the polygon on the left

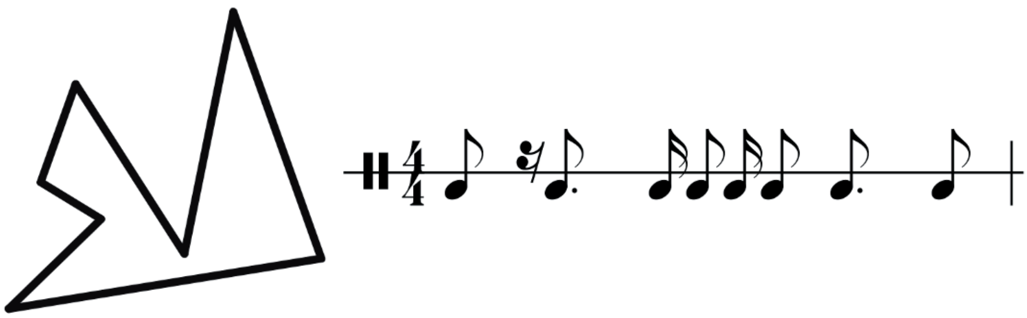

Figure 9: One possible rhythm for the shape on the left

The Output Layer outputs the rhythms generated and chosen by the Composition Layer.

4. Evaluation

4.1 Characterization

For the evaluation of our methods we created two web questionnaires. Both were composed by personal questions (user age, music skills, music instrument skills, painting interest and painting knowledge) and by 8 questions. In each of the questions, we asked the users to evaluate a music track, in a 0 to 6 scale, following a painting. In the first questionnaire, the music was generated by our system, and in the second one, the music was generated randomly.

The questionnaires were distributed electronically by email and social networks. The targeted users belonged to no special music or computer research groups, and there were a total of 47 participants in the first questionnaire and 35 in the second one.

The questions we were trying to answer with this test were:

- Is the music generated by Sense2 appropriate to the paintings?
- Does music knowledge and skills influence the participants' opinion on the music's appropriateness?
- Does painting interest and knowledge influence the participants' opinion on the music's appropriateness?

4.2 Music Appropriateness

Figure 10 presents a comparison between the average appropriateness from both tests: normal and random.

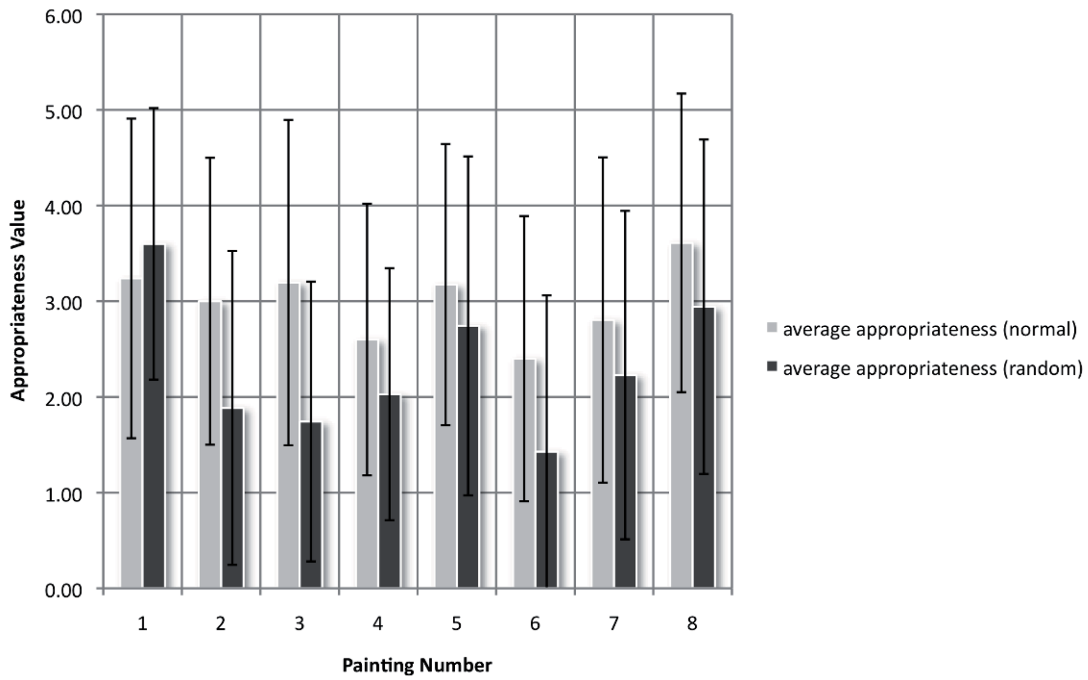

Figure 10: Comparison of average appropriateness value per painting in the normal test and random test

The correlation factors between the average appropriateness from normal generated music and all the personal characteristics can be seen in table 2.

	average appropriateness
music knowledge	-0,15
music skills	-0,24
painting interest	0.29
painting knowledge	0.15
global painting (mean between painting interest and knowledge)	0.33

Table 2: Correlation factor (ρ) between the average appropriateness from normal generated music and personal characteristics

To identify significant differences between both average appropriateness measures (from music generated normally and randomly) we performed the Mann-Whitney U tests. These results are presented in table 3.

	average appropriateness
Mann-Whitney U	425,500
Asymp. Sig. (2 tailed)	0,000

Table 3: Mann-Whitney U differences test

In the first test, it appears that people saw some similarity between the painting and its music. The standard deviation values are close to 2, which is very high and means that the appropriateness varied quite a lot between users and that there is no consensus between users in their opinion. The value 3 is typically a value used by subjects who are undecided, so having an average of 3 shows perhaps that the participants did not know what to answer, so they answered in the middle of the scale.

However, the correlation factor between the painting characteristics and the average appropriateness is moderate and not insignificant, specially the global painting one, meaning that the higher the painting knowledge and interest people have, the higher relation between the painting and its music they see. The correlation factor between the music knowledge and the average appropriateness is negative but too weak to conclude about its influence on appropriateness. We should be careful when talking about the correlation factor between music instrument skills and average appropriateness because we only have two percussion player responses to our forms, and the correlation does not mirror that.

The comparison between the values of music appropriateness of both tests shows that there is a clear difference between the appropriateness of both tests. The Mann-Whitney U test confirms this, as the 2-tailed value is much lower than the p-value of 0.0.5. This means that there is not enough confidence to accept the null hypothesis, i.e. the results obtained in the randomly generated music test are significantly worse than the results obtained in the normally generated music test.

Form 2 always ‘looses’ to form 1 except in the first painting, but the average difference is d = 0.68, which shows that appropriateness from form 2 is a bit lower than the appropriateness from form 1. The average difference for standard deviation for appropriateness of both forms is d = -0.024, which means that the results from form 2 are insignificantly more disperse than the results from form 1.

5. Conclusions & Future Work

5.1 Conclusions

At this time, the questions which emerge are: did we manage to create a system which provides a synesthetic experience? And did the generated music transmit the same feeling as the painting? The answers to these questions are not straight and the coin is always two-sided.

The results from the evaluation show that the system had some impact, but that this one varied from person to person. Some users admitted that most of the music tracks did not make sense to them, whereas others said that the system really made sense to them. First, we think that the fact that the domain is too abstract, too vast and the possibilities too many, may have contributed to some wrong choices. And second, we think that the users’ bar was too high, and that the users were expecting a perfect system, because they do not have a notion of the difficulty of the area.

5.2 Future Work

Given the evaluation results, we propose future work to improve our model. Some users referred colors as being important to their perception of the painting. Even if colors do not have a direct influence on instability or stability perception by the user, they could be present in the ambient or the melody of the music. Dark colors mixed with few shapes transmit a dark and slow ambience and maybe a more frighten or suspense feeling. Whereas dark colors mixed with lots of confusing shapes transmit a dark but fast ambience, maybe like an anguish feeling. Light colors transmit a more happy and agitated ambience. Music can easily transmit that. Color interpretation mixed with shapes interpretation is a whole new field and could be a good improvement to our matter.

Also, these models could be used in other intermedia projects. We believe that instead of shapes, we could use the shape instability model in dance movements, in a dance-music relation, resulting in the 'sonification' of live dance pieces.

References

[1] Richard Cytowic. The Man Who Tasted Shapes. MIT Press, 2003.

[2] Edward Hubbard and Vilayanur Ramachandran. "Hearing Colors, Tasting Shapes." Scientific American, April 2003.

[3] Daniel Smilek, Kelly Malcolmson, Jonathan Carriere, Meghan Eller, Donna Kwan, and Michael Reynold. "When "32" is a Jerk and "E" is a King: Personifying Inanimate Objects in Synesthesia." Journal of Cognitive Neuroscience 19, no. 6 (June 2007): 981-992.

[4] Hajo Duchting. Paul Klee: Painting Music. Prestel Verlag, 1997.

[5] Wassily Kandinsky. Concerning the Spiritual in Art. Dover Publications, 1977.

[6] Myra Melford. "Aural Architecture." In Arcana: Musicians on Music, by John Zorn, 119-135. New York: Granary Books, 1999.

[7] Hannah Higgins. Fluxus experience. University of California Press, 2002.

[8] Eduardo Lopes. "Just in Time: Towards a theory of rhythm and metre." Phd Thesis, Department of Music, University of Southampton, Southampton, 2003.

[9] Eduardo Lopes. "From Blues To Latin Just In Time: A rhythmic analysis of 'Unit Seven'." Jazz Research Journal 2, no. 1 (2008): 55-82.

[10] David Cope. Computer Models of Musical Creativity. The MIT Press, 2005.

[11] Stephan Schwanauer. “A learning machine for tonal composition.” In Machine models of music, 511-532. Cambridge: MIT Press, 1992.

[12] David Cope. The Algorithmic Composer. A-R Editions, 2000.

[13] Peter Copley and Andrew Gartland-Jones. “Musical form and algorithmic solutions.” C&C '05: Proceedings of the 5th conference on Creativity & cognition. London, 2005. 226-231.

[14] George Papadopoulos and Geraint Wiggins. “AI Methods for Algorithmic Composition: A Survey, A Critical View, and Future Prospects.” Proceedings of the AISB'99 Symposium on Musical Creativity. AISB, 1999. 110-117.

[15] Stephen Pope. “Fifteen Years of Computer-assisted Composition.” Proceedings of the 2nd Brazilian Symposium on Computer Music. 1995.

[16] Ricardo Cruz. “I-Sounds: Emotion-based Music Composition for Virtual Environments.” Msc Thesis, Instituto Superior Técnico, Lisbon, 2008.

[17] Ricardo Cruz and Eduardo Lopes. “Pulse Salience Quantification Method.” Document obtained from the authors, 2008.

[18] André Gonçalves. Pulse Salience and Instability calculation. 08 May 2009. http://web.ist.utl.pt/andre.pintado/salience (accessed November 01, 2009).