From Movement to Music, A Computational Creativity Approach

Edgar Carita Miguéns Santos Meireles

Thesis to obtain the Master of Science Degree in

Information Systems and Computer Engineering

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

Examination Committee

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

January 2021

Acknowledgments

First of all, I would like to thank my supervisors, Prof. Helena Sofia Pinto and Prof. Nuno Manuel Robalo Correia, whose guidance, knowledge and patience, helped me pursue this goal of mine.

Secondly, I would like to thank Prof. Rui Pereira Jorge, whose support and share of musical knowledge helped the development of this work, where otherwise would not have been possible.

Thirdly, I want to thank my parents, Carla and José, and my grandparents Maria Conceição, João, Maria Teresa and Fernando for all the support and strength given during this period, without whom I would not have been able to complete this work.

Lastly, I would like to express my gratitude to all my friends who supported me, and with whom I shared my doubts. A special thank you for my dear friends Francisco Centeno, Ana valadares, João Dias, Carla Martins and Luís Aleixo.

Abstract

Creativity in itself is a hard to define human only ability. The creative process, is then, something researchers aim to fully define and recreate in artificial ways. In our work we go through the historical and cultural definitions of creativity, and some theories of the creative process.

Computational Creativity is a field of Artificial intelligence that seeks to create a program able of creative thinking, as well as creative acting.

In this document we present an attempt of making a creative system capable of human body motion capturing and music generation based on what is captured. We present our approach to this problem with a cross domain analogy between the visual to the musical domain.

We ended making three different motion to music approaches and evaluate them further along in this thesis.

Questionnaires were made for this evaluation. The results proved to be good as to evaluate if we matched our goals proposed on this thesis. Our system was considered creative and able to generate music by more than 72% of our evaluators. We also confirmed an association between the movement and the music with over 56% of common adjectives when categorizing the movement and the music.

There are an almost infinite different ways to solve the initial problem. Our proposition is nevertheless, in our point of view, an interesting and successful approach.

Keywords

Computational Creativity, Cross Domain Analogy, Computer Vision, Computational Music Composition

Resumo

Criatividade em si é uma habilidade humana dificil de definir. O processo criativo é então, algo que os investigadores almejam conseguir definir e recriar artificialmente. No nosso trabalho nós percorremos as definições de criatividade e do processo criativo ao longo dos tempos e em diferentes culturas.

Criatividade Computacional é uma sub-área de Inteligencia Artificial, que tem como objectivo criar um programa capaz de pensar e agir de forma considerada criativa.

Neste documento nós apresentamos uma forma the produzir um sistem capaz de captar o movimento do corpo human e gerar musica baseado no que foi captado. Nós apresentamos a nossa abordagem a este problema baseado em analagoias entre os domínios visual e musical.

Nós acabamos por desenvolver três associações diferentes de movimento para musica e avaliamo-las mais a frente nesta tese.

Questionarios foram usados para esta avaliação. Os resultados provaram-se bons a quando avaliamos se cumprimos os nossos objetivos propostos nesta tese. O sistema foi considerado creativo e capaz de produzir musica por mais de 72% dos avaliadores. Foi também confirmado uma associação entre o movimento e a musica gerada atravéz de adjectivos comuns dados a interpretação do movimento e da musica, estes sendo mais de 56%.

Há quase infinitas formas diferentes de resolver o problema inicial. Mas apesar disso, nós acreditamos que a nossa proposta é uma abordagem interessante e de sucesso.

Palavras Chave

Creatividade Computacional, Analogia entre Domínios, Processamento de Imagem e Visão, Composição Musical Computacional

Contents

List of Figures

List of Tables

List of Algorithms

1

Introduction

Contents

The work we present belongs to the field of computational creativity. Computational creativity is a field that studies means to make a computer program act in a way that is considered creative, or produce something that can be considered creative. There are many historical events in which humanity shows the capability to be creative and in the 21st century can be an enduring and survival skill [30]. This capability is one of the characteristics that allowed humans to evolve and make artifacts that helped us get on the top of the food chain, improve the way we live, and, is one of the abilities that makes us constantly create new things, as explained by Margaret Boden in [5].

Creative problem solving has lead humanity to what we are today and led us into making things that were, in some moment in history, unimaginable.

Computational creativity may help mankind to solve problems that we currently have, or even optimize already found solutions. If we introduce this ability into machines, we could perhaps reach a whole new level.

1.1 Movement and Music connection

Every species has a way of mobility and its own way of moving. Movement can be a way for one to express itself. As humans our body language can express what we are feeling, and we invented dance as an performing art using movement with our bodies to express ourselves. Movement is not only used by humans to express one, irrational animals are also found to do so. A courtship display is seen in the wild life were usually a male performs a dance to a female as attempts to attract a mate.

Music is also a form of art, this one captures our auditory sensations. Music is found in every corner of our globe, from the most tribal groups to the most modern societies, now and in the past.

There is not and accepted connection theory between music processing and dance association for human beings. Darwin proposed that music and dance might have evolved for these courtship displays as a species needed to find a way to select a better mate [19].

But there is indeed a connection between both of these art forms. As these are art forms they are generally associated with creativity. As we studied the subject we found interesting to develop a system capable of doing the inverse of this association. Trying fist to process the movement or dance and generate music with it.

1.2 Problem introduction

This thesis propose a way to develop a creative system able to process videos of people moving and generate music associated with it. Our main problem to solve is so this movement to music process.

A moving person in a two dimensional visual state as it is on a video, to be translated into a music domain piece is a complex problem. Using divide and conquer, we divided the problem in sub problems. Processing the video is our number one problem, as we need to be able to translate what we see and need a data sequence for further analyses. Than we need to explain what we see for us to be able to express what we see into something else. With this we can now make a association between what we see and what we want to hear. Finally we need to be able to produce our final music piece with the association made before.

Looking at our sub problems all of them have their challenges. For us the more challenging one is the third as this one have a large variety of associations depending on what we want to extract from our video. So we found useful to also divide this problem into two. We divided it into a primary association and secondary association. The first problem is for us to make a more superficial association and based only on some specific features. The secondary is for us to produce a more complex association based on a larger number of features. This division made sense since we also aim to study how these associations are received by our evaluators.

Now looking at each sub problem individually, first we have to be able to perceive what we see. There is a large number of visual features is the visual domain perceived by a human being. From color to lines, from shapes to shades. There are two main compounds as we see a video: background and foreground. Here we propose that these two features are key to our problem. Also color is a good feature to work on, we see colors, and color is the thing that defines every other visual feature we can calculate.

For us to explain what we see we need to translate what we see into something that can be evaluated. If we want to explain what we saw in words adjectives are the best way, in computer science numbers is the way to go. So we need to translate the features that we want to extract into numbers for us to process them and explain them in a computer science way.

To transform what we see in what we hear we studied first the way of music composition in music theory and then using what we see we need to be able to translate it into what are we supposed to hear.

Once we figured out what are we supposed to hear when we see x, we need to be able to produce our association into a music piece. when solving this final problem we are able to solve our main problem and produce music based on the movement of a person seen in a video.

1.3 Motivation and Objectives

We aim to develop of a program capable of producing creative products as well as showing creative procedure as making them. Humanity was always inspired by sounds, as shown in the creation of music instruments. There are many studies that show that music can have a important role in our day-to-day life, can improve our health, our mood and even our decision making [3, 23, 26].

Some of the movement we humans make are also inspired by the sounds or music we are listening.

People tend to walk according to the rhythm of what they are hearing or thinking about [36]. The idea of our project was to make music inspired by the movement of a certain person, or group of persons, since normally the inspiration or creative flow comes in the other way around, as seen in dance for instance, a dancer usually creates its choreography based on music. We thought that it would be very interesting to study this inverse flow of creativity, and to make a computer program capable of it.

Our main objective is for the program to be inspired by a set of features extracted from human motion, and, with that, generate a sound that map the movement executed by the human from whom the program is inspired by and for that sound to be considered music.

The second objective is for the music created to match the movement from which the program that created it was inspired by. This would mean that not only could the program detect the motion rhythm and intent, but also capture it and be creative enough to create a piece of music that can translate a movement into a different field.

As our final objectives we want to be able to produce a system that is considered creative.

Our motivation is to contribute to multiple fields of study such as Computational Creativity, Computer Vision and Computational Music Composition with a cross-domain approach. We hope to build a system capable of that task, giving the user a much more pleasing sensations with both visual and auditory stimulus than when perceiving only one of them.

So our main goal is for the program to be able to capture what we see into what we listen and both of them to make sense, be considered music and a creative object when combined and perceived by our audience that later evaluate our system.

1.4 Document Structure

In chapter 2 we describe the previous work in the fields of Computational Creativity, Human Body Motion Recognition and Computational Music Composition. Then, in section 2.4, we explain and compare the models and techniques used on the previous sections and provide an overall overview on the related word in section 2.5.

In chapter 3 we talk about the visual domain processed in our project. We divided this chapter in three sections, color in which we analyse what is color, background and silhouette are for the explanation of the analyses we made of them in our project

Chapter 4 explains us the musical background necessary for the implementation of our system.

The implementation specification, architecture, as well as drooped attempts when we were developing our system are described in chapter 5.

Later, chapter 6 is where we analyse and evaluate our system based on our goals, inputs and results obtained.

Finally chapter 7 concludes this document with some insides of what was this project as well as future ideas of how to continue the development of it.

2

Related Work

Contents

A program able to act in a way considered creative is a task that has been keeping many researchers busy for a long time. Since we aim to build a system that can be considered creative, and to make a translation of human body motion to music composition we see three different areas of study worth gaining some knowledge upon: Creativity, Computational Creativity, Human Body Motion Tracking and Recognition and Computational Music Composition. So, before we can start to develop a solution to our problem, we first need to see what has been made so far in those fields.

2.1 Creativity

2.1.1 Creativity Background

For our project be considered as one in the field of computational creativity, first, we have to understand creativity and define it. Then we need to comprehend the creative process and how to evaluate if a certain object is considered creative or not.

The word ”creativity” comes from the Latin word ”creare” which mean ”to create”. However creativity was only defined later in history. R Keith Sawer in [31] say that in the ancient Greece to be creative meant to have been possessed by a demon, as a divine gift granted by the gods only given to certain selected individuals. It was also thought that creativity came from some sort of mental illness, and, through times there were created some myths about creativity like ”Creativity comes from the unconscious”, ”Creativity Represents the Inner Spirit of the Individual” and ”Creativity Is Spontaneous Inspiration” among many others. Although, as we study creativity, we learn that there is a certain procedure in order to produce something creative. So, creativity can never come only from the unconscious since it needs conscious hard work in figuring the problem out, thus can not also be spontaneous inspiration.

Margaret Boden defines creativity as ”the ability to generate novel, and valuable, ideas.” [5] This is one of the most accepted definitions by the scientific community. In this case, a valuable idea can be seen as one that is interesting, useful, beautiful or even extremely complex. Basically everything that has some sort of purpose in a field can be seen as valuable. From paintings to algorithms, from complex sculptures to a simple photograph, any idea, physical or not, can be considered to be creative as long as that idea is new and valuable. Boden divides novelty into two categories:

- **P-creativity**

  Is defined as psychological novelty. This means that the product is a new idea for the individual, it does not matter if another person has already tough the same thing before in history.

- **H-creativity**

  H-creativity is a P-creative idea that has never been though through history. Something completely new.

Creativity is perceived from various aspects, a product or a process can be considered to be creative by some cultures but not others, can even be considered by only some people from a certain culture but others may find it banal and not creative at all. Creativity depends on the act itself, on the person who makes it, on the people who perspective it and on the product itself.

What is left for us to see is what is a creative product, what can be considered creative or not.

Taking Margaret Boden's definition in consideration, anything that is new and valuable can be considered creative, it does not say that the product has to be something physical as our intuition tells us. A idea, a theory, even a simple joke can be considered to be creative.

2.1.2 The Creative Process

R Keith Sawyer, an American psychologist who has study creativity in his book *Explaining Creativity: the science of human innovation* [31], defines the creative process in four basic stages:

- **Preparation**
  No one can be creative without first being a specialist in the area they are working on. This is the stage where the person gets the data about the field, learns the techniques necessary for the correct execution of the tasks needed in the field, trains and perfects those techniques to be an expert. This is the most time consuming of all four stages of creative process. The next stages can not occur without this one.

- **Incubation**
  When a person finds a problem that he aims to solve, he will use every piece of information he has on the field in order to solve it. Incubation is when the person absorbs and organizes the knowledge gotten previously. This stage mostly happens unconsciously as the person arranges and rearranges his thoughts to produce a solution.

- **Insight**
  This stage is the ”eureka” stage. Previously the person got the data, became an expert, and, when facing a problem tries to use every resource he has available in order to solve it. Insight happens when the person suddenly thinks of a possible solution to the problem he was struggling with. Psychology has yet to come with a explanation to how and why Insight happens. There is no proven theory, for the extension of our knowledge, that can say what is the thing, or things, that trigger this stage on a person.

- **Verification**
  Now that the creator has a presumed solution, he has to evaluate the solution he came across during the Insight stage. This final stage of the creative process is divided in two sub-stages:

> ”Evaluation” and ”Elaboration”. In Evaluation the person finds itself wondering if the idea is in fact viable, interesting or just a theory that might work. In Elaboration, when finally the idea has been mentally made viable and possible, the creator starts to work on it and idealize the solution to the problem.

This definition can help us to create a program capable of creative execution since the inspiration comes from human creativity.

2.2 Computational Creativity

In [5] Margaret Boden explains that there are three ways to produce something novel and valuable, making it creative according to her previous statement. It is said that novel ideas may come from combination, exploration or transformation.

Combinational creativity is when it is combined ideas never linked before. Combining ideas from different fields is mostly assumed that this is the only way to be creative. However there are two more ways of being creative worth mentioning.

In exploratory ideas, the person sees the constrained rules of a certain space of thought and explore it to its maximum. Making things still on that domain, and still limited by its constraints, but, yet new and never seen before.

Transformational creativity happens when the person is familiar with the space and its limitations, and expands it dropping one of the limitations. Pablo Picasso on creating cubism, saw the limitations of human representation in paintings and dropped the overall rule of his time to create something new.

On the contrary of one may think, Combinational Creativity is, of the three, the most difficult to replicate on a computer program. The computer can in fact combine ideas of different fields easily but it is fairly difficult for it to join them in a interesting way, therefore make that combination valuable to the field.

Margaret Boden talks about the work of Binsted, Pain, and Ritchie in 1997 on making JAPE (Joke Analyses and Production Engine) [4]. As said, this is one of the best models of computational Combinational Creativity. JAPE's jokes are based on combination of word meanings, and, had a studied structure for its jokes.

Jape's joke production model has two main components, the first is the retrieval of words information, and the other is the joke structuring and making. With the combination of words compounds, and the theory of humor this system was able to produce something we actually think is creative.

Exploratory Creativity can also be modeled into a computer program. To model it one need a lot insight of the field and the boundaries that limit it. It is not needed a lot of expertise in Artificial intelligence to achieve this type of Computational Creativity. There are many systems developed with this model of creativity presented by the author in the fields of music, architecture, visual art and more.

People tend to say a program cannot achieve Transformational Creativity since a program is defined by its execution rules, therefore it can not be able to achieve this type of creativity since it cannot overcome these rules. But if a program could change its execution steps along the way, then it transforms its domain space thus achieving Transformational Creativity. This can be made with evolution algorithms that through mutations and crossover of features can generate new code for the program to transform itself.

2.3 Human Body Motion Tracking and Recognition

Human body motion recognition and tracking is a field of computing vision that has been studied for a long time.

In 1996 [15] there was a proposed solution for modeling the human body in three dimensions using multiple cameras. With multiple views of the body in motion it was possible to adjust the it into six cylinders, one for the whole body, one for the torso, and one for each member.This way they could make the representation of the whole body of the humans.

Once they had the body segmented, in order to keep track of the body motion, they implemented a framework based on Joseph O'rourke and Norman I Badler work [28]. This framework consists in four different components and they work cyclically. The first component is the Prediction component where they try to predict the body model state in the next frame based on the speed and direction each component is taking in the previous frames. Then comes the synthesis component that models the prediction of the body in the next frame given by the previous component. The third component is Image Analysis where they remove the background and get only the body frame of the person they are trying to model. Then comes the State Estimation component when based on their three dimension human body model they position each individual component of the body they model. In this last state they make the adjustments between the prediction and the actual body position then comes the prediction state and so on frame by frame.

Another approach was later developed by Yang Song, et al. in their paper called *Unsupervised Learning of Human Motion* [34]. Their aim was to develop a program capable on recognizing weak models of human body. Those models are the ones where physical geometric body models such as the ones we previously analysed are not possible to compute, or when the outcomes have lots of errors. This models are used when the person we aim to detect does not have their body well shown, for instance, when a person is wearing a set of clothes that are more loose. In their work they modelled the human body by the joint probability density of their position and speed of the body parts. Their work

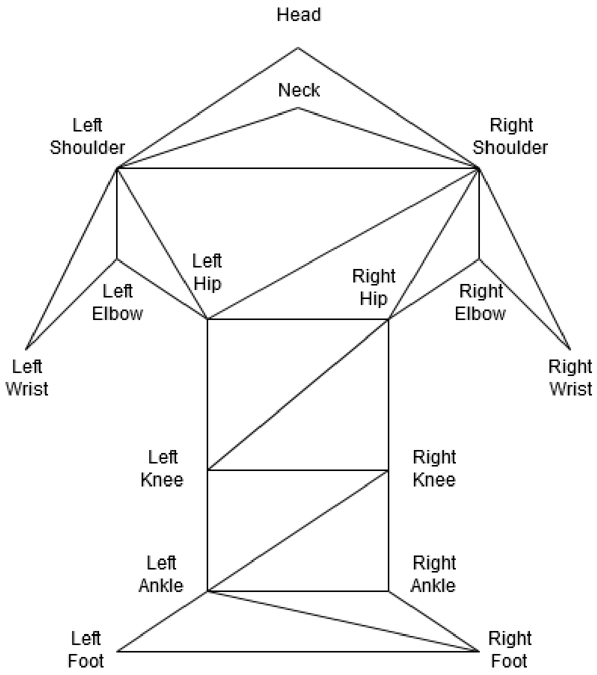


Figure 2.1: Triangular graph of Human Body

was able to identify keyjoints of the human body without labelling them, and based on their position, speed, keyjoints position approximation probabilities and on a human body triangular graph model they developed, shown in figure 2.1, it was possible for the program they developed to learn through a set of frames the whole posture of the person.

In 2005 Claudio Fanti et al. [12] used the triangular graph human body models used in the previous presented paper to recognize human motion through Hybrid Models. So they improved the previous approach using hybrid Models for learning the human posture in a video. This Hybrid probabilistic model combines multiple variables such as whole body position, and local quantities in the body joints like velocities and appearances. With this approach they could in fact improve the velocity the previous system learned and improved the body joints and posture recognition.

Continuing the same idea Cristian Sminchisesc et al. in 2006 [33] tried to improve previous approaches of unsupervised learning of human posture using triangular graphs this time with Conditional models.

Many more techniques have been studied since than. Now we have pieces of software based on the previous explored work that gives us the human body model and labels of joints. This made possible the approaches we are now presenting, since all of them already use automatic human body detection.

The work of Achyuta Aich et al. on [1] was quite different than previous approaches. In their work they captured correct Bharatanatyam dance postures from a dance instructor using kinect cameras to get the stick figure, as shown in figure 2.2, then, using those analysed postures, they could develop a

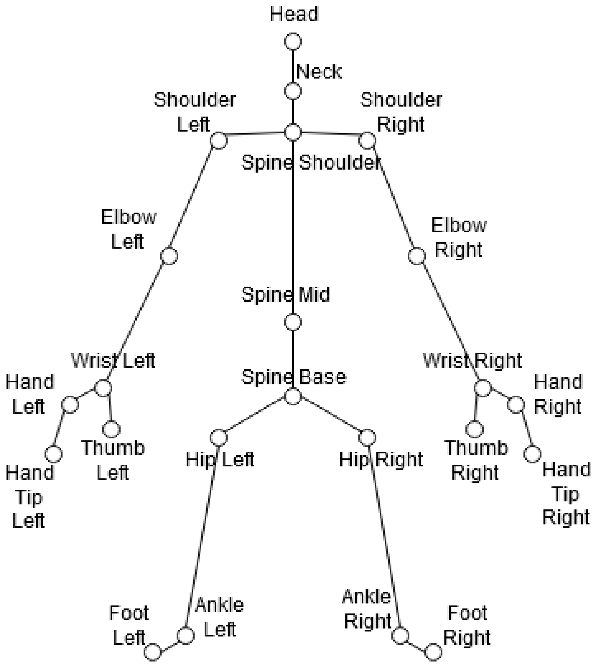


Figure 2.2: Kinect stick figure of human body

program capable of correcting a posture of Bharatanatyam dance apprentice through the stick figures similarities, adjusting the apprentice posture to the correct one.

Later Tanwi Mallick et al. on [24] used the previous approach of human body motion recognition to recognize Bharatanatyam dance choreographies. Their program receives a video, and frame by frame they see similarities with the postures of the Bharatanatyam dance expert. With this approach they could get the choreography of that dance with the time/frame constant.

Until now we talked about human posture and keyjoints detection. Since we aim to translate body movements to sound and music we have to not only detect and track the human body but also translate the movement we perceive into a set of features. Velocity and acceleration are features that we can detect with the keyjoint detection, but if we can aim to correctly translate the movement into sound we need more features.

Then we came across the work of Emma Frid et al. [14]. The authors of this paper studied the relation between movement fluidity and the type of sounds that we humans associate with them. They found out that a fluid movement is associated with less frequency notes as well as lower bandwidth, on the contrary non fluid movements tend to be perceived and translated as high frequency and higher bandwidth notes. They say that the sound to better translate a fluid movement is continuous calm and slow, sounds like water or wind came to mind to the users they experimented on. When we talk about non fluidity the best signification is a hash, non-continuous sound, dissonant or robotic sounds were also associated with this kind of body movement.

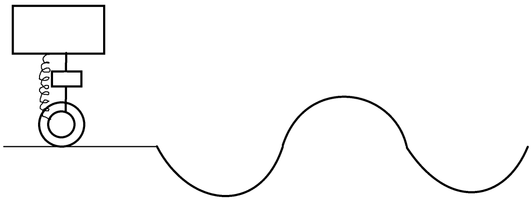


Figure 2.3: Mass-Spring-Damper model

Taking in consideration the previously analysed paper, fluidity can, in fact, be a good feature to extract from body movement, so when investigating more about fluidity we found the work of Stefano Piana et al. [29]. In this paper they propose a definition in body motion fluidity based on Mass-Spring-Damper model [10] (see figure 2.3). The idea is to think of body joints as masses and strings and if their movement behave as Mass-Spring-Damper model suggests they should, then the movement is to be considered as fluid.

This theory has been tested and used in some studies like [16, 27]. They used this theory to see athletes effort during runs and jumps. If a movement is non fluid than the body is in a high level of effort since the body stop acting as a natural mass would. This proves to be a good feature to extract, since we can also get the effort that the performer we are captioning is putting in their movement.

Later we thought that since music can transmit certain emotions, then maybe we can try to make emotional analysis of human motion.

We found a paper entitled Emotional Body Movements [32] which gave us some useful theories about how to extract emotion from motion. The authors start to explain that the Laban Movement analyses is a set of guidelines to consider in order to create a series of movements to create particular effects on the spectators. They say that there are four major components to detect emotion according to the Laban Movement analyses which are Body, Space, Shape and Effort. Body describes the body parts that move, Space describes where the body is moving and its amplitude, Shape is how the body moves, and Effort is the combination of Time and Flow of the movement. The emotion that we humans capture from motion comes from the combination of this four aspects.

Some years later Kenji Amaya et all. developed a model to produce emotional animations in [2]. Their method is based only on two of the features we talked about before that are speed and space. With the analyses of real human motion doing a set of different tasks each of them made with different emotion attached, the authors of this paper could reproduce the emotion in animations with the impersonation of the general speed and spacial amplitude used by the persons they studied on.

The work of Ginevra Castellano et al. on their paper *Recognising Human Emotions from Body Movement and Gesture Dynamics* [7] show that in fact the four components we previously talked about could

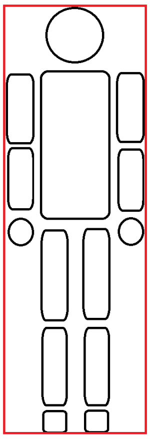


Figure 2.4: Smaller bounding box

in fact be used to detect emotion. They used five features of motion detection: Speed, Acceleration, Fluidity, Contraction Index and Quantity of Motion. Speed and Acceleration are extracted by the key joint position in space. Fluidity was extracted by their definition which said that a movement is fluid if in a movement from two specific points the general acceleration feature remains 0 or close to 0. Contraction index is a value that comes from the division of the area of the person and the smaller bounding box that could fit that person (rectangle that can fit the person body inside it like in figure 2.4). A person is more contracted as the contraction value increases.

The last feature was Quantity of motion(QoM) that is given by the following expressions where SMI stands for silhouette motion image and a Silhouette is an image of a person removing the background, in this case, Silhouette value is the area or number of pixels the person corresponds to:

$$SMI[t] = (\sum_{i=0}^{n} Silhouette[t-i]) - Silhouette[t]$$

$$QoM = Area(SMI[t,n])/Area(Silhouette[t])$$

So by the formulas we see that SMI in a number of frames is the sum of the variation from frame to frame of the Silhouette area value. And QoM is the SMI value over a number of frames divided by the silhouette value in the frame we are analysing.

With these features they developed a automatic emotion detector from motion for the four basic emotions, Anger, Joy, Pleasure and Sadness using K-nearest-neighbours, decision trees and Bayesian Networks as classifier algorithms.

2.4 Computational Music Composition

Artificial intelligence has been in the field of computational music composition for quite some time now.

Jose D Fernández and Francisco Vico in [13], summarize many of the possible artificial intelligence approaches to music composition. They say that algorithmic composition can be grouped in four categories:

- **Symbolic Artificial Intelligence**

  Under this group we have rule based systems which have been proven quite effective since they can learn and reason over a set of rules given by experts in the field. An example of music composition with this kind of approach is given in 2008 by Georg Boenn et al. in [6]. Using the melodic composing rules of music, the authors were able to create a system capable of music composition called ANTON. Although this system can produce melodic pieces, it cannot do entire pieces of music.

- **Machine learning (Markov Chains and artificial neural networks)**

  Most machine learning algorithms tend to imitate the input or categorize it based on what type of data the model is trained.

  Music can be viewed as a very complex and sophisticated probability distribution over a sequence of sounds. With this premise, many researchers adapt Markov models and artificial neural networks to be able to learn this probabilistic sequences in order to make new sequences of notes, that later could be called music.

  In 1957 Lejaren Hiller and Leonard Isaacson composed the first known computer made piece of music. This was a string quartet composition made with pseudo random Markov chains [21]. The notes the code generated were later tested with a certain number of rules and they only kept the ones that agreed with all the rules they implemented. This is a basic probabilistic approach to the music composition problem. With the evolution of AI we see another approach to this problem, we see many people trying to make automated music composition using machine learning algorithms.

  In 1992 Hermannn Hild et al. used Artificial Neural Networks in [20] to learn Bach composition rules to make compositions similar to his style.

  Also in 1992 we see another proposition with Artificial Neural Networks [25]. Here the authors propose that recurrent neural networks, a specific type of Artificial Neural Networks, can learn music structures, though be able to compose music with higher quality. So Florian Colombo et al. in [9] used recurrent neural network to learn music structures of Irish melodies. With the structure learning capabilities of this algorithms, the music composed by them has better and more similar structures to when using normal Artificial Neural Networks.

Long Short-Term Memory, a different type of Artificial Neural networks, is also very used to music composition in [8,11]. On these papers, the authors claim that although Recurrent Neural Networks can in fact learn music structures, Long Short-Term Memory algorithms have better result in timing and context of music structures.

- **Optimization techniques (evolutionary algorithms)**

  Darwin told us that we evolve as species through generations as well as all other species on the planet. Based on that theory, this type of algorithms were created. Given a population as input, this algorithms creates new generations with the good features of the older generations. These good features are define by a fitness function given by the programmer.

  In music composition we can use these types of algorithms are used to combine a group of musics to make new ones. With a good adjustment of the fitness function we can see good, never seen or listened results, from this approach.

- **Self-similarity**

  The authors say that these techniques are not a form of artificial intelligence. This method consists in using similarities or musical patterns and repeat them to compose. Usually the music composed by these systems are very rough yet they are certainly novel. This is used mostly by composers to create raw material for them to compose on.

In 2017 Joana Teixeira made a system capable of producing music inspired by images. [37] Features were retrieved from an image, and, with a visual to music features association made by the author, the program was able to directly translate what it sees to music.

2.5 Models & Techniques

Now that we have shown some of the work developed in the fields in study, we now aboard the models that fit the most with our proposal.

So we analyse in more detail the kinect camera approach used in [2,7]. On those papers, the authors used keyjoint detection and with them extracted features to analyse emotion in human body motion. This approach has much more adaptability since we can use it with a simple video without the need to create a large data set. With this model we can also extract the features we aim to get from the human body motion, such as velocity, acceleration, fluidity, quantity of motion and contraction index.

With this in mind we further analysed the software we had available for keyjoint detection and we found two alternatives. The kinect sensor that is talked and used in some of the previous approached papers, and a tensorflow package named posenet. A kinect sensor can detect between one and four

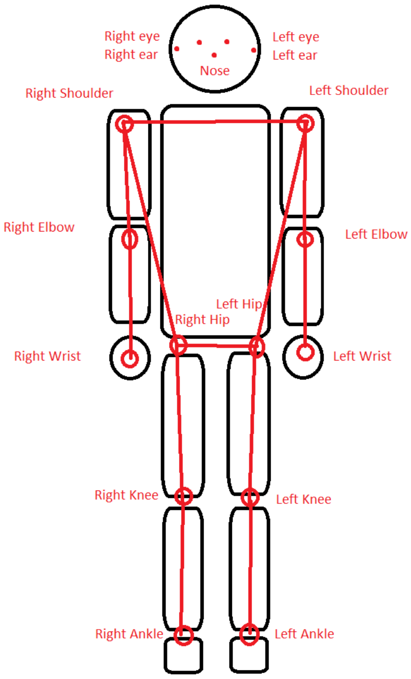


Figure 2.5: tensorflow posenet package keyjoints detection

bodies, the tensorflow package can detect as many as we want. However the tensorflow package can only detect 17 human body keyjoints (see figure 2.5) when the kinect sensor can detect 25 keyjoints.

In Computational Music Composition we see four major models explained in [13] and another on approached in [37].

Symbolic Artificial Intelligence are quite good in the structure part of music composition. Although they involve a lot of previous knowledge in music which give us the need to see and acquire that knowledge through experts. These systems are rule based ones. When the knowledge is acquired from experts we then build a model with feature dependencies. As said before we can create good music structures and good note dependencies thus possibly creating music with overall good quality.

As said in [17], grammars are good for long range dependencies like seen in music structures, thus making them a good approach form music composition.

Machine learning algorithms on the contrary do not need that much music knowledge from our part in order to implement them, but, they need a large data set of musical pieces and they might become biased into a certain music style, the one we would use to train these models.

The most commonly used models for music composition through machine learning are Markov Models and Artificial Neural Networks as seen on the previous section.

Markov Models are stochastic models used for randomly state changing systems. In these models it is assumed that the next state directly depends on the current one. In figure 2.6 we have a example of a

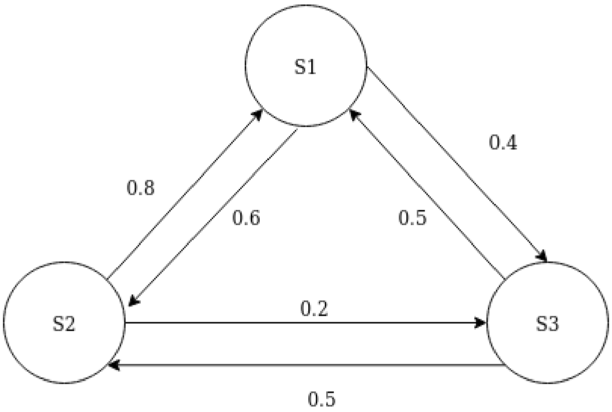


Figure 2.6: Markov Model

simple Markov Model. There we see that if we are in State ”S1” we have 60% chance to go go to change to state ”S2” and 40% to go to ”S3”. To use them in music composition we only need to associate each note to a state and give a probability from the next note based on the consonance of those two notes together. So, by randomly changing states in these models we get a sequence of notes that is a music piece.

Artificial Neural Networks, on the other hand, are algorithms that when given a training set of data divided in features, the algorithm train itself by weighting the features and its combinations.

Figure 2.7 is an example of a Artificial Neural Network. The input layer there shown in that case is a group of four features, than for every single neuron, the balls in each hidden layer, it adds up the features, multiplied by the connection weights, and with that value calculate another one with an activation function. With this final number the neuron passes it to every node of the next hidden layer.

When every neuron of the final hidden layer has its final value, we add them up and then we get our output value. All this process is called forward propagation.

For This algorithm to learn, we then compare the value with the real expected output for the inserted input values, calculate the value and update the weights of every connection between neurons, inputs and output. This process is called Backward propagation.

Repeating this process a certain number of times, the algorithm gradually weights its parameters correctly and so be able to output much more alike what is expected with the input it receives.

This is the algorithm for basic Artificial Neural networks. Recurrent Neural Networks are talked in some of the work we saw is a type of Artificial Neural Networks that has some backwards connections. This is when a neuron in a more advanced hidden layer send its final value to a neuron in a sooner to be calculated hidden layer.

We can apply these algorithms to learn with multiple songs we find to be music good quality. Given the right musical features, these systems are able to sequentially compose good music based on the features it captured from the learnt musics.

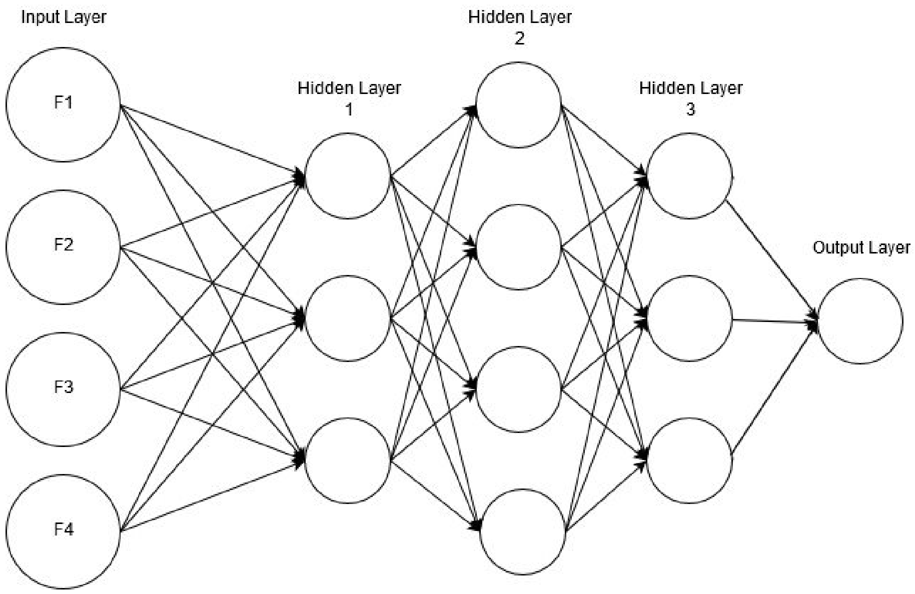


Figure 2.7: Artificial Neural network

Optimization techniques also do not need for a very large knowledge coming from us and they do not need a music data set as large as the machine learning approach. This model although effective in creating novelty through mutations is not appropriate for our cross domain approach since our initial point would be a video and not a music for it to mutate. Yet it can be use full as a mapping technique between the video and music domains. Genetic Algorithms are the most common algorithm used by this approach.

Genetic algorithms are algorithms that create new populations based on the progenitors (initial population). As seen in the figure 2.8, genetic algorithms are composed by three major components that run cyclically a number of times the programmer decides: Fitness Function, Selection and Reproduction.

The Fitness Function is what determines if a certain part of the population is good or bad with the features previously determined by the person who will implement the code. The Selection is where the algorithm selects the good individuals of the population based on the Fitness Function. And finally, the Reproduction phase is where new population is generated based on the good parts of the selected individuals.

Self-similarity as said by the authors of [13] does not compose music that is aesthetically appealing. It is mainly used for raw material and to be later worked over as a base.

Finally the last model we saw was through inspiration and cross domain analogy. This approach is the closest in terms of structure and way to approach the problem to our aimed work, as far our knowledge. The author used a cross domain analogy for a direct translation from an image to music. It was also used optimization techniques in the model proposed in [37] for composition of multiple midi files as output. We call this approach ”Direct Mapping” in this document, since it mapped directly the

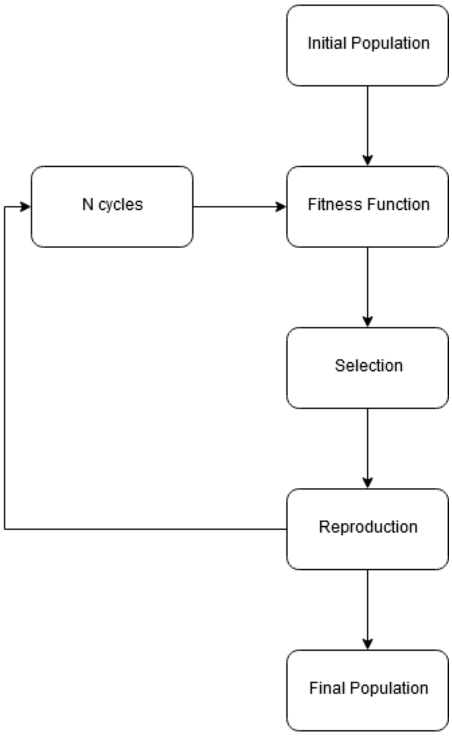


Figure 2.8: Genetic Algorithm model

Table 2.1: Human body motion tracking and recognition overview, KS:Kinect sensor, TPP:tensorflow posenet package

Model/Technique	#people detected	motion detection	#keyjoints detected
posture matching	1	not capable	not capable
KS keyjoints tracking	1-4	capable	25
TPP keyjoints tracking	not limited	capable	17

features from one domain to another.

2.6 Overview

As a overview, we now compare the techniques approached on the previous sections and justify our choices of implementation talked in section 5.

We approached three models of computational creativity. Among the three Combinatorial Creativity is the objective model for our system.

We aim to build a system able to combine ideas from human body movement and music in order to produce a product considered creative, thus novel and valuable in the field of music composition.

The cross domain analogy is a requirement in our system so, of the three models Combinatorial Creativity is the one to go.

From the three alternatives (seen in table 2.1) figure we see that posture matching is not very good

Table 2.2: Computational Music Composition overview

Model/Technique	need music knowledge	need music database	music quality
Symbolic Artificial Intelligence	+++	-	++
Machine Learning	++	+++	++
Optimization Techniques	++	-	++
Self-similarity	+	-	-
Direct Mapping	++	-	++

on the characteristics we aim. This leaves us with the trade-off between number of keyjoints detected and number of people the model can detect. We want to build a system capable of produce sound and music with most of the videos we give it, so, we choose the tensorflow posenet package to detect human bodies since we can detect more people and we do not need all the extra keyjoints kinect sensor detects.

In detecting Fluidity we see two models that do so [7, 29], the first one is based on the Mass-Spring-Damper model and the other one is based on the overall movement acceleration variation. Between the two we see that the acceleration variation method is simpler and gives overall good definition of fluidity. The approach based on the Mass-Spring-Damper model would give us more specific results and better insight over the fluidity of those keyjoints. This approach however is worth as to give overall movement fluidity values and is much harder to compute. Therefore acceleration variation method was the approach we chose to use.

In table 2.2 we analyse the models that we talked about for the field of Computational Music Composition. Since our time is short the Symbolic Artificial Intelligence is not an option for us.

Machine learning have the need for a search of a large database to train the models however the optimization techniques, self-similarity and inspiration based models do not need them.

Now for our final three options, we exclude Self-similarity models since we want our output as good as possible in terms of music quality. Between Inspiration models and optimization techniques we decided to choose inspiration models since our group has experience in these kind of systems and decided to inspire ourselves in optimization techniques for the implementation merger, explained further along in this document.

3

Visual Domain

Contents

A video in computer science can be seen as an array of frames, these frames are defined by a 2D array of pixels, then finally a pixel can be define with three doubles. The position of the pixel on the 2D array is, so, its position on the frame.

For us human beings processing all of the images we capture with our eyes is natural. We perceive colors, shapes, space, texture, all of this is perceived and processed unconscious and instantly as we go on with our lives.

A human perceive a video the same way we perceive everything that is captured with our eyes. What we see is delivered to our brain as instant images and are processed sequentially. This gives us the notion of motion.

We can instantly tell what is moving and what is static from this sequence of images. Our attention is drowned to what is moving, to that object or set of objects that moves in space while the rest of the image does not change.

While processing all of this is challenging enough task, as a human being capable of seeing, we also perceive and retrieve emotions from everything that is happening around us. Certain colors can influence our mood [22]. Certain body languages can tell us what people are thinking and feeling [38].

As a computer all of this needs to be translated to a computer in order for us to be able to develop our system.

3.1 Color

Color is everywhere and in every little thing. There are infinite numbers that we perceive as human beings. It is something that surrounds everywhere and every time and has such a strong influence in our lives. Color perception differs from person to person, it is a subjective thing. Our culture, our mood, the context. Colors are associated with many things. Many cultures use various colors to specific emotions, or assign other symbolic meaning to them. In some English speaking cultures, red can be associated with anger, green with happy and so on as seen in figure 3.1.

Color Theory can so describe us the logic and usefulness of colors. Color theory describes color as a combination of three properties: hue, value and saturation. Hue tells us from the color wheel which color we are seeing from the twelve hue categories from the color wheel.

The color wheel was originally divided in three colors, the primary colors: red, yellow and blue**??**. In traditional Color theory, these are the colors that can not be obtained when mixing other colors, and all colors are derived from them. Then if we divide the wheel in six we also see the secondary colors: Green, orange and blue. These colors appear from the equal mixture of two primary colors. For instance, red plus yellow gives us orange. Tertiary colors appear when we divide the wheel in twelve equal pieces. These colors appear from the combination of primary and secondary colors: Yellow-green,

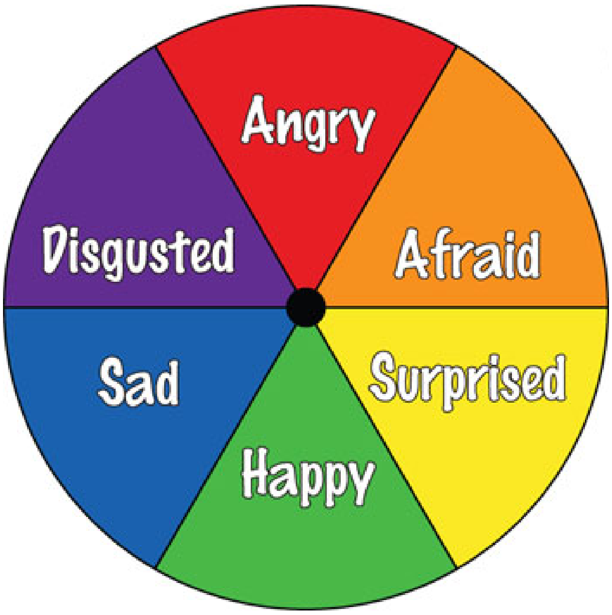


Figure 3.1: Emotion color association

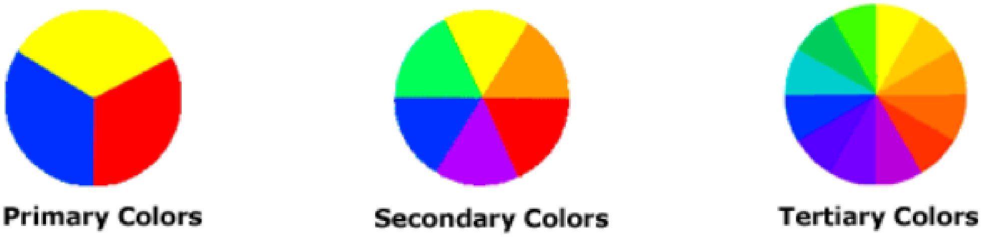


Figure 3.2: Primary, Secondary and Tertiary Color Wheels

yellow-orange, red-orange, red-purple, blue-green and blue-purple.

All these color Wheels can be seen in figure 3.2

We could divide the hues into even more colors but Color theory says that larger than twelve hues would be near impossible to see all of them as clear different colors, and also it would be harder to memorize over 20 different relevant color hues. From these twelve hues categories or colors. We can draw a line in the middle of them that separates warmer colors to colder ones. 3.3

In our color wheel, there are two colors that we humans can clearly see and distinguish from the others: black and white.To identify these colors come the second and third properties. The second property, value, it defines how light or dark that color is. There are infinite red colors in the hue category red. And we can see some differences among some of them. If we increase the value to its maximum, the color will reach the maximum lightness thus becoming white. Otherwise if we decrease the value to its minimum, we can no longer see the lightness of the color red for instance, thus becoming pure black.

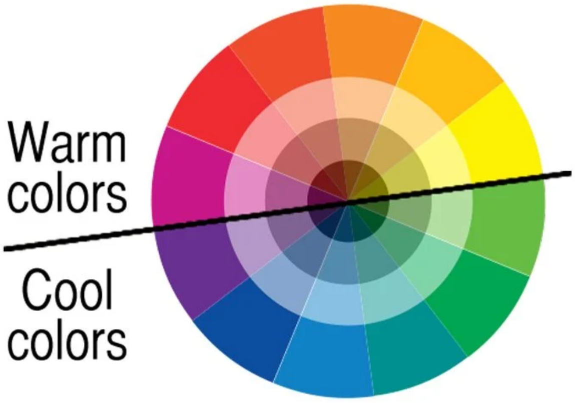


Figure 3.3: Warm and cool colors in the color wheel

Finally the third color property, intensity, tells us how pure the color is. This compares a dull greyer red with a bright red.

3.1.1 Color Models

In computer science there were defined many models that can describe color. In our work, we focused on two diferent models, RGB (Red Green Blue) and HSV (Hue Saturation Value)

- **RGB** RGB is an additive color model that defines all its colors as a sum of its three color components red green and blue.

  This color model is defined by a tuple of three integers in the range of 0 and 255. It can so represent over 16 million different colors.

  This model is better visualized in three dimensions, where x,y and z coordinates represent the red, green and blue values respectively. Since our space is limited from 0 to 255. The color space is a cube as seen in figure 3.4

- **HSV** This Color model defines color also as a three numbers, yet this model defines them in a different way than the previous one.

  HSV comes from the name of the three components Hue, Saturation and Value. Hue is the same as Color Theory describes it. In this model is defined as an angle from 0º to 359º that gives us the color. Value is defined as a double from 0 to 100. It is also the same mentioned in Color theory yet when it reaches its highest value the color does not become white but yet it becomes the brighter version of that hue. Saturation is also defined as a double from 0 to 100 and gives us how much grey there is in a color.

  This color model can be seen in a three dimensional space as a cylinder.But since when it reaches

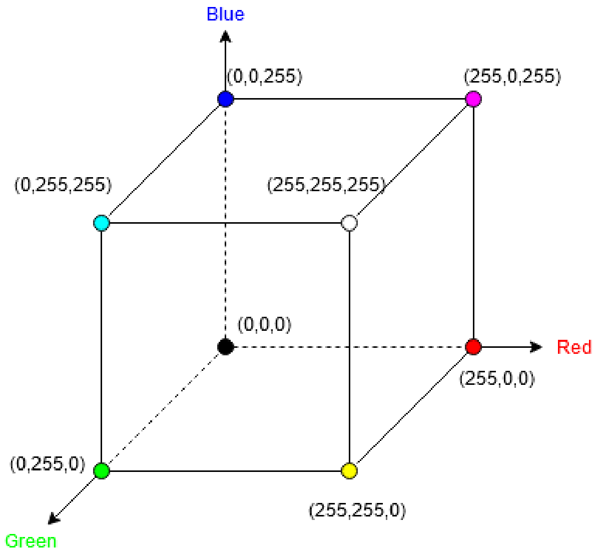


Figure 3.4: RGB color model

the lowest value it becomes black regardless of the other two components. So illustrators usually represent it as cone as seen in figure 3.5

To switch from RGB model to the HSV we just follow the following formulas with the R, G and B values:

$$R' = \frac{R}{255}$$

$$G' = \frac{G}{255}$$

$$B' = \frac{B}{255}$$

$$Cmax = max(R', G', B')$$

$$Cmin = min(R', G', B')$$

$$delta = Cmax - Cmin$$

$$H = \begin{cases} 0, & delta = 0 \\ 60 \times (\frac{G'-B'}{delta} mod 6), & Cmax = R' \\ 60 \times (\frac{B'-R'}{delta} + 2), & Cmax = G' \\ 60 \times (\frac{R'-G'}{delta} + 4), & Cmax = B' \end{cases}$$

$$S = \begin{cases} 0, & Cmax = 0 \\ \frac{delta}{Cmax}, & Cmax \neq 0 \end{cases}$$

$$V = Cmax$$

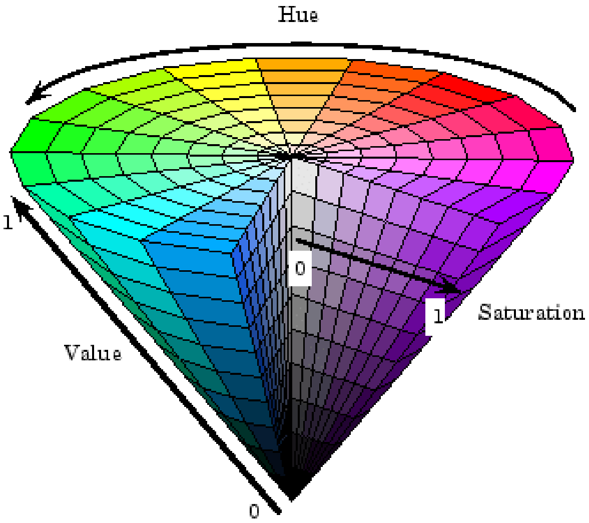


Figure 3.5: HSV color model

3.2 Image Background

Foreground detection is one of the major tasks in the field of computer vision. The ability to detect changes on a sequence of frames is something trivial for us but for a computer is something that can be very challenging. Background subtraction is one technique that allows the foreground to be analysed specifically by subtracting the background to the original frame.

This technique is highly used for the detection of moving objects from static cameras. The most commonly approach to implement this technique, is to first use a temporal average filter to identify the background to then later subtract it to the main image. This filter, with a given number of previous images, estimates the median value of every pixel. Since the moving object does not remain in the same place this technique is able to remove it with the median values thus removing it from the background detected.

Then to recognize the moving object, as said before, it is just needed a subtraction of the obtained background to the following frames and we get the foreground for the next frames. This technique can be used in static environments with a static camera, this way the background does not change so the initial calculation of the background can be used for the rest of the video. If the background changes a bit, like a chair is moved from one place to another, this technique considers the chair as moving object. To avoid this error we need to continuously perform a background detection using n previous frames from the one we are analysing.

As said, an image is an agglomerate of a number of features. To our project we found relevant to calculate two features: main color and main lines.

An image can have a number of different colors equal to its resolution. In this project we only use

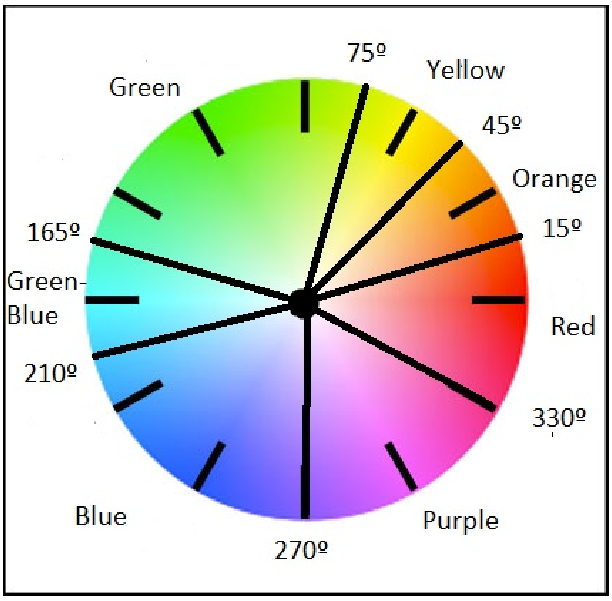


Figure 3.6: Seven rainbow colors in the hue clock

Table 3.1: Examples of image filtering

Name of the filter	Usage
Frangi Vesselness filter	Used to see likeliness of an image region to contain vessels
Meijering neuriteness filter	Used for detection of ridges eg. rivers.
Sato tubeness filter	Used for detection of ridges eg. rivers
Gabor filter	Used for image texture analyses
Gaussian filter	Used to blur images, reduce image noise
Prewitt transform	Used for image border detection

seven hues 3.6 corresponding to the seven colors on the rainbow plus black and white since it was simpler for later to associate them with music features.

To get the main lines of the background we had to use some techniques present in digital image processing [18].

Digital image processing is the use of algorithms to process an image with a computer. Image is considered as a multi-variable function and with the use of mathematical transformations we can blur and or sharpen it.

A digital filter is a system that performs mathematical operations on an signal, to reduce or enhance certain aspects of it. It is characterized by its transformation function or difference equation. If Images can be seen as distorted by noise. So image filtering basically is the filtering of it. Filters re-evaluate the value of each pixel in an image.

We found these as some examples of image filters relevant for our project, table 3.1.

Background detection and image analyses is fundamental for our project.

3.3 Silhouette

The person silhouette is what is considered to be the foreground in our project since we have a static camera and background.

Once we have the silhouette we thought about the characteristics we want to get from it.

When we watch a person moving, the movement is of course the main attraction of our visual stimulation. But we also see the colors the person is wearing. Darker color may give us different feelings then when a person is doing the same movements wearing whiter colors. So we also calculated the main color of the silhouette using the same approach as we did on the background.

For the movement we want to be able to perceive it as translate it into a language our program can understand. So we want for the program to be able to perceive the person keyjoints for every frame. Giving it the X and y coordinates on it. From there, the person is therefore translated into a set of points with specific coordinates for every frame. With that we can use math to calculate a large numbers of features that we humans can perceive naturally.

As we read on related work made in this area, chapter 2, velocity, acceleration, fluidity, quantity of motion, contraction index, emotion are good features to extract from a person movement. These features are the main focus of our work regarding our visual domain interpretation.

4

Musical Domain

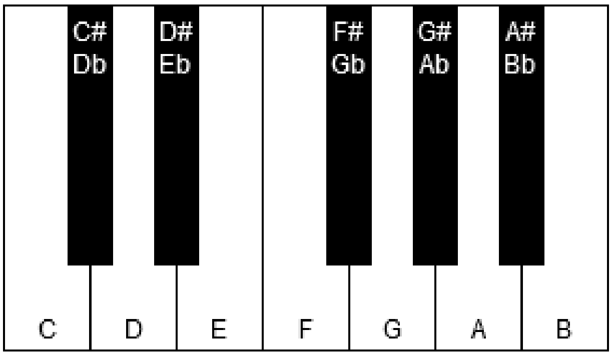


Figure 4.1: Piano keys

Music is a form of art and cultural activity that captures our auditory sensations. Music is a combination of sequences of sounds and silences, that as an art form can express sensations and is individual the opinion of one about weather he likes it or not.

A music has a piece has a beginning and an end, and each sound has its own characteristics. Music production is not made at random, sounds and silences are put together in that sequence with logic. This logic comes from music theory.

Music theory is the study of the infinite possibilities and practices of music. It gives us the basics and fundamentals one should have learned in order to produce any music piece.

A music piece is often written in music sheets. This is a form form of musical notation that uses music symbols to indicate some of the sound characteristics. As said, sounds are present in music, these sounds are called notes. Notes have many features, the main ones are Pitch, Duration, Intensity and Timbre.

- **Pitch**:

  Pitch is the frequency of the sound wave that comes from a note, usually measured in hertz, number of cycles per second. This frequency allows us to rank each note on a scale. For instance if one sound/note has a higher frequency we can put it further away in the scale as when compared with a lower frequency one.

  As a tuning standard, it was decided that the frequency 440Hz would be the standard tune for the middle A note.

  From this standard there were created twelve pitch categories, or notes, this is called the twelve tone scale. These categories are designated with the alphabet letters from A to G.

  Figure 4.1 shows us a keyboard visualization of the twelve notes.

  As seen the twelve notes can not be characterized just with the sequence from A to G since there are only seven letters on this sequence. As in figure 4.1 there are five more notes that have one of the letters with a ”#” or ”b” next to eat.

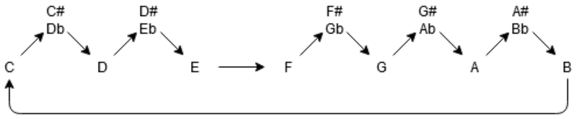


Figure 4.2: Music notes

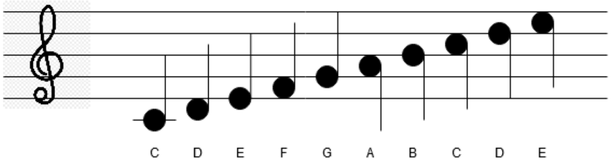


Figure 4.3: Music Sheet

The ”#” symbol means that this note is half tone higher than the note its letter represents. A# for instance is half tone higher than the note A. This note is called a A sharp.

The ”b” symbol means the opposite of the symbol ”#”. It means that this note is half tone lower than the note its letter represents. Ab is half note lower than the note A. This note is called a A flat.

After these notes it all repeats itself, a half tone higher than every B is always another C, but this C is an octave higher than the previous one. As well as half tone before every C there is always a B. This B is an octave lower than the previous one.

An octave is the most perfect note interval there is when played simultaneously. It is characterized as a six tone note intervals and for every six tones there is always the same pitch category or note as before.

In figure 4.2 we see a visual demonstration of the notes cycle, every note that is another each other is at a half tone distance from that note.

A blank music sheet is a five horizontal line combination repeated on a blank piece of paper. On these five lines we can draw the notes based on its location on theses five lines and on the clef the music is written on. The clef is a musical symbol used to indicate where a certain note is on the five lines.

Notes can be represented on the lines or between the lines, every time you go up a semi line you go up on the scale. Figure 4.3 shows a visual implementation of notes displayed on a musical sheet with a G clef.

As seen on figure 4.3 the note is define with a ball that is called notehead and a stick that is called a stem. Some notes are displayed with a curved line from the top of the stem, that line is called a flag. To know which note is represented with that symbol we see where the notehead is and, for instance, the G note is where the G clef symbol starts.

sharp and flat notes are displayed on the music sheets with a ”#” or a ”b” before the notehead.

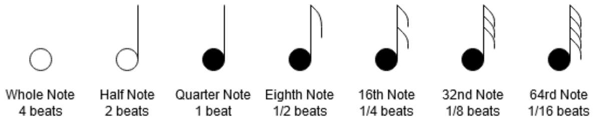


Figure 4.4: Notes Duration

- **Duration**

  Duration is how long that note is going to sound.

  As one learns music, he learns some symbols that appear on a music sheet that represent the duration of that particular note. Figure 4.4 shows the notation used.

  As seen in figure 4.4 the duration of a specific note comes in beats.To translate this value into seconds, the person who writes the musical piece defines a value that is called bpm, beats per minute, from there we just have to use the following formula.

$$time(s) = \frac{\#beats \times 60}{bpm}$$

- **Intensity**

  Intensity is basically how load a note is played. This refers to the amplitude of the sound wave that note produces

- **Timber**

  Timber is the perceived sound quality of the note played. Timber is the characteristic that is able to distinguish weather a certain note is played by a certain musical or was played by a choir or an individual voice.

In a music sheet there are also other elements worth mentioning. As said before, we have a bpm value, usually displayed on the top left corned where we start our music sheet. This value is called the tempo of a music.

Through the advances of musical theory some names have been given to some bpm values. Table 4.1.

Also in the beginning of the music sheet, right next to the clef we see a fraction like 4/4 or 3/4. This is called the time signature. This fraction is used to define how many notes fit in a staff. Staff is a vertical line that divides the music sheet in parts.

The bottom number present in that fraction, typically a power of 2, indicates the note value that represents a beat. For instance every fraction of the type n/4 says that every beat is represented by a quarter note. The upper number of the fraction represents how many beats we have in a staff. A 4/4 time signature says that for every staff we can fit up to four quarter notes in whatever combination possible.

Table 4.1: Tempo terms by bpm values

Tempo	BPM intervals
Larghissimo	under 25
Grave	25-45
Largo	40-60
Lento	45-60
Larghetto	60-66
Adagio	66-76
Adagietto	72-80
Andante	76-108
Andantino	80-108
Marcia moderato	83-85
Andante moderato	92-112
Moderato	108-120
Allegretto	112-120
Allegro moderato	116-120
Allegro	120-156
Vivace	156-176
Allegrissimo	172-176
Presto	168-200
Prestissimo	over 200

Table 4.2: Types of scales

Type	number of notes per octave	usage
Chromatic	12	sequence with every existing note
Octatonic	8	used in jazz and modern classical music
Heptatonic	7	most common modern western scales
Hexatonic	6	common in western folk music
Pentatonic	5	used in folk music and Asian music

Notes are sometimes put together in groups that are called scales. One can compose a music piece just with the notes of a scale.

Scales are defined as a succession of notes, this succession is made with the notes distance from the next or previous one. This distance is half tone (H) for every consecutive notes and whole tone (W) for notes when you skip the next note of the twelve existing ones. For instance the distance between C and D is a whole tone since the next note after C is C# and only then it comes D.

There are some types of scales as seen in 4.2

There are also tetratonic, tritonic and ditonic scales that were dropped, they were only used in primitive music. Among the Heptatonic scales there are two that are the most commonly used:

- **(M)ajor Scale:** The succession of it is the following: W - W - H - W - W - W - H.

  When starting in C we have the most common major scale, C Major (C - D - E - F - G - A - B). This scale defines the white keys in a piano, the black keys do not belong to this scale.

- **(m)inor Scale:** The succession of this scale is: W - H - W - W - H - W - W.

> Note that A minor scale has the same notes as C Major scale. This means that A is the minor relative key of C major scale. To obtain the minor relative key of a major scale you have to go one and a half tones before the tone of your major scale.

Among the pentatonic scale types we also have a major scale that is widely used. This scale has a succession: W - W - (W + H) - W - (W + H).

The C Major pentatonic scale has the following notes (C D E G A). Note that all of the notes in this scale belong to the corresponding heptatonic major scale of the same note removing the fourth and the seventh notes.

As a last major feature music has its structure. This one is usually defined by the genre rules in which the music composed is inserted in. All of these rules are on the topic of music theory. For one to compose, it has to be familiarized with all of these rules and restrictions.

To the human being music is what sounds and feels right when listened to. It is a form of art capable of expressing feelings and emotions. A person can listen to a musical piece and in a matter of seconds tell weather it likes it or not, and weather if it is music or not.

One of the main goals of our system is to develop a system capable of music generation. To perform such task we had to first familiarize ourselves with music theory for later to translate it into a language understandable by a machine.

5

Implementation

Contents

The focus of this chapter is the project implementation as well as the choices made while we were developing the project.

We chose python as our programming language since we have a long experience of working with it and this language is very flexible. We used python 3.7.4 since it was the most recent version as when we started this project.

Furthermore it has a vast number of packages that we need for video and image processing as well as processing data.

5.1 System Architecture

For this project we divide the main project in five small problems. First we want to process the video, then we want to get our visual characteristics that come from the video. Once we get our visual features we want to associate them with musical features. The fourth problem is to create a new musical association not made only by the visual features direct translation. Then we want to be able to generate a midi file with the music characteristics we generated.

With this five sub-problems, it made sense to create an architecture with five modules, each module created to solve one of the five sub-problems. 5.1:

- **Video interpreter**
  This modules purpose is to receive a video as input and return an array of keyjoint positions per frame.

- **Feature extractor**
  This module was implemented to extract the visual features we studied. It receives the keyjoint position array and the video as well. It returns

- **Feature associator**
  Once we get the visual features this module creates music features associated with the features created on the previous module.

- **Implementation merger**
  Now with the music features, this module mixes some of previous module associations in order to avoid direct translation.

- **Music Composer**
  Finally this module generate midi files with the music features generated on the previous modules.

The input video is pass to both Video processor module and Feature extractor module as both needed them to solve their problems. Video processor is the first module to run and after their procedures it returns the keijoint positions per frame to the Feature extractor module. Then Feature extractor calculates the features mentioned on chapter 3 and passes them to Feature associator. Feature associator gives the music features it calculates to not only music composer but also Implementation merger, for this last module to compose with the previous calculated music features. Music composer is then going to receive both implementations made by feature associator and one extra implementation made by the Implementation merger. With this the music composer is going to output three .midi files, one for each implementation generated by the previous modules.

5.2 Video processing

This modules purpose is to get the human body keyjoint position per frame.

To do so we used python tensorflow pose net package. This package is able to give the position of each keyjoint previously mentioned with a certain confidence.

We used this package algorithm with single pose feature on, so that the algorithm only search for one person per frame. We used the smallest output stride and the highest image stride so that we get the better results with this algorithm. This results in slower processing but since we do not aim for real time processing as one of our goals we chose this approach.

So we ran this algorithm for every frame of the video and got an array of the type [id, x, y, confidence] per keyjoint, being the id the keyjoint id, the x and y are the coordinates of the keyjoint on the frame, and finally the confidence is the algorithm confidence that that is the keyjoint position on the frame. We then create a bigger with all of the keyjoint array on that frame.

We then return an array of those frame keyjoint position arrays.

5.3 Feature extractor

Now with the processed video, we then calculate all the features we previously mentioned, using the keyjoints captured on the previous module that have a confidence percentage value over than 0.5:

Velocity (*v*), can easily calculated with the previous keyjoints coordinates using the following formula for every 2 sequential frames:

$$v = \frac{\sum_{i=1}^{k} \sqrt{(x_{i2} - x_{i1})^2 - (y_{i2} - y_{i1})^2}}{k}$$

Being *k* the number of keyjoints, *xiN* is the x coordinate of keyjoint i on the nth frame and *yin* is the x coordinate of keyjoint i on the nth frame.

Acceleration (*a*) can be calculated from velocity by the expression:

$$a = v_2 - v_1$$

Being *vN*, the velocity on the frame N. Frame 1 and 2 are sequential.

Fluidity (*f*) is, so, calculated from acceleration with a similar formula:

$$f = |a_2 - a_1|$$

Being *aN*, the acceleration on the frame N. frame 1 and 2 have to be sequential.

Here we assume that a fluid movement is one that has a low acceleration change rate.

The contraction index (*ci*) is obtained from the area of the bounding box. We use the code of the algorithm 5.1 to get the silhouette bounding box per frame. An example can be seen in figure 5.2.

Algorithm 5.1: Calculate Bounding Box limits

```
maxX ← -inf
maxY ← -inf
minX ← inf
minY ← inf
for all keyjoint i: keyjoints do
  if i[x] > maxX then
    maxX ← i[x]
  end if
  if i[x] < minX then
    minX ← i[x]
  end if
  if i[x] > maxY then
    maxY ← i[y]
  end if
  if i[y] < minY then
    minY ← i[y]
  end if
end for
```

And then we apply the following formula to get the contraction index value:

$$ci = (maxX - minX) * (maxY - minY)$$

Quantity of movement (*qom*) is the next feature we extracted. This one was calculated based on the variation of the contraction index value. It was calculated using the following formula:

$$qom = ci_2 - ci_1$$

Being *ciN* the contraction index value on the frame N. Also with frame 1 and 2 sequential.

We found useful also to calculate the variation of quantity of motion (*2ndqom*). So we calculate it based on the previous obtained quantity of motion value with the following formula:

$$2ndqom = qom_2 - qom_1$$

Being *qomN* the contraction index value on the frame N. As always with frame 1 and 2 sequential.

From these features we create an array of the type [v,a,f,ci,qom,2qom] per frame starting of the third frame. This happens since acceleration is calculated with the values of velocity of the current and the previous frame. If N is the total number of frames we have, then we can only get N - 1 acceleration values. And since fluidity is calculated with the same formula but with acceleration values, we only have N - 2 fluidity values.

We also have a second array of the type [vup, aup, fup, vdown, adown, fdown]. We created this as a second type of visual feature extraction.

From the formula, velocity is calculated as the average velocity from a certain number of keyjoints. For the first array we use every eligible keyjoint (confidence value over 0.5) for our velocity calculation.

On the second array we have 6 ”new” features. These features come from the analyses of the keyjoints from above the waist and from the waist and below. So vup, aup, fup are velocity, acceleration and fluidity calculated with the keyjoints above the waist. vdown, adown,fdown are so, the velocity, acceleration and fluidity captured from the keyjoints from the waist and below.

As we saw the videos we adjust two threshold values of contraction index and quantity of motion for the program to be able to perceive the emotion transmitted with the captures movement. These values were generated based on our personal perception of emotion transmitted from those movements.

With these values we created a 2 dimension chart from which our program is able to identify emotions per frame as seen in figure 5.3.

Joy was associated with the value 0, pleasure with 1, anger 2 and sadness with the value 3. We then count the instances of these values and pass to the next module the most common emotion value associated with the video body movement.

After this we also extracted the main color of the silhouette and the background.

To do so we first needed to see as the first frame per instance what was background and what was foreground, in our project the foreground is the person silhouette.

For the background identification we used the median per pixel to see the most common color it appears on that pixel. This works since the silhouette is always moving through the whole video so it does not appear on the median color for every pixel.

Now we have the background we remove it from the first frame to obtain the silhouette. To do this we go through every pixel of the background and every corresponding pixel on the first frame and if the RGB values difference is lower than 20 we replace that pixel with the color black.

With these two images we go through every pixel and label it with one of nine color, seven rainbow colors plus black and white, as mentioned in chapter 3. To do so we translate the RGB values to HSV. For the seven rainbow colors we used the hue values as represented on figure3.6. The color black is seen as any color with the value below 0.3. For white we say that every color with saturation under 0.15 is considered white. Then we count those colors and we get the most common color on each image.

Now for the last visual feature we also wanted to see how much of the background can pop into a user eyes as visual stimulation.

To do this we use the background image we obtained previously and apply a Gaussian filter to blur it, since our peripheral vision also does it.

Now in the blurred background image, color contrasts may call our eye to it so we apply now a Meijering neuriteness filter to transform our blurred background into a black and white image. The white pixels represent pixels identified as ridges, or in our case a visual stimulation since it represents a contrast of the background.

Meijering neuriteness filter is used since it was the one after that after many tests that prove to be the best one for our project, with our group visual perspective opinions.

By counting the number of white pixels and divide them with the total number of pixels, we get a percentage value of how much of that picture is attractive to our eye and can disturb our focus on the main attraction.

Both the Meijering neuriteness filter and the Gaussian filter used were given by the python package skimage.filters.

We use that value as our final visual feature, we called it background attraction value.

5.4 Feature Associator

From the previous module we received a number of features:

- Array of overall motion [v, a, f, ci, qom, 2qom] per frame
- Array of specific parts motion [vup, aup, fup, vdown, adown, fdown] per frame
- Overall emotion

- Background main color
- Silhouette main color
- Background attraction value

The last feature is not used in this module, we just passed it here to pass it to the next module.

In this module we created an algorithm that could pick three continuous features and one fixed value and from that create an array of notes for the composer module to generate midi files.

First of all we decided that every music has two tracks, so we should run the algorithm twice to get our two tracks, each with different visual features.

For each track we assigned a musical instrument. These instruments come from the fourth and fifth features given to us by our previous module.

Based on a personal choice, we associated a color to an instrument:

- White - Electric Guitar
- Black - Sax
- Blue - Piano
- Green blue - Cello
- Green - Violin
- Yellow - Harp
- Orange - Flute
- Red - Celesta
- Purple - Guitar

We used the silhouette main color to get the first music track instrument. The background main color is used then to get the second track instrument.

Now for the main algorithm we start to generate a random number from 0 to 11. This number represents the main note of the musical scale the generated music composes on.

The emotion value defines the type of scale our system composes on:

- Joy - Major scale
- Pleasure - Penthatonic Major scale
- Anger - Chromatic scale

- Sadness - minor scale

Then we generate the scale based on the main note with the type previously defined.

Since midi notes range goes from number 0 to 127, and most instruments can not play that range we decided to limit our note range from 57 to 92. This range correspond to A3 to G#6, so we work on three octaves. We choose this range to fit every instrument we may generate on.

So when we generate both the main note of the scale and the scale itself, we run an algorithm that returns an array of numbers that correspond to the notes of the scale.

This algorithm starts with the first possible note on the scale and adds the note intervals specific of the scale we want.

For instance for the A major scale our algorithm returns the array:

[**57**, 59, 61, 62, 64, 66, 68, **69**, 71, 73, 74, 76, 78, 80, **81**, 83, 85, 86, 88, 90 ,92]

Since we start with the A3 that corresponds to the number 57 and from there we add the note differences correspondent with the major scale, in bold are the three A notes present on the scale.

Now with the array of notes that belong to the musical scale we can start to generate the music.

Firstly we needed to find a way to discover the tempo of the music we wanted to generate. In other words, what is the value of beats per minute our music has.

To find the tempo of the video we observed the fluidity values as a wave. A wave have a periodicity. So the number of frames associated with the periodicity of our wave is the number of frames associated with a beat.

If we think of a person walking, and think of every time a person puts his foot down, there is going to be a break in the legs acceleration since it stops moving.

So we decided to apply a Fast Fourier Transform on the derivative of acceleration, in our case the fluidity value, to obtain the Discrete Fourier Transform of our wave.

Extracting the maximum of our Discrete Fourier Transform gives us the most likely periodicity of our wave in number of frames.

To get the beats per minute value (*bpm*) we used the following formula:

$$bpm = \frac{frameRate * 60}{max(dtf)}$$

Being *dtf* the values of our Discrete Fourier Transform.

We decided to take in consideration the tempo range we studied and generate our music in the range between 60 and 120 bpm since is the middle range of the studied tempo ranges. We found that values outside these values did not match the movement. Whenever our Discrete Fourier Transform maximum was not in this range we extracted the next maximum value that fited our range.

Now it did not make sense to see our visual data in terms of which frame x happened. So we reduced our feature array by grouping them in groups of the number of frames presented in a beat, then we average all the values presented in that group to get all the features we previously had but now in beats instead of frames.

As mentioned before a note has four features: pitch, duration and intensity and beat.

In our project we defined intensity as a constant value. We used the maximum value possible in midi, 100, because as our time was limited we did not thought of how or which visual features could be ”translated” into this intensity value.

Beat is defined as the time the features we analyse occurred.

Pitch is defined as one of the notes presented in our pre-generated scale, and as duration, both are chosen by the algorithms 5.2, 5.3 and 5.4:

Algorithm 5.2: Note generator

```
notes ← []
noteIdx ← int(random() $\times$ size(scale))
timeBeat ← 0
while timeBeat $<$ size(featuresByBeat) do
  beatFeatures ← featuresByBeat[timeBeat]
  transaction ← getTransaction( beatFeatures[1], beatFeatures[2] )
  noteDuration ← getnoteDuration ( beatFeatures[0] )
  if noteDuration $<$ 1 then
    nBeats ← 1
  else
    nBeasts ← noteDuration
  end if
  i ← 0
  while i $<$ nBeats do
    noteIdx ← noteIdx + transaction
    if noteIdx $<$ 0 or $>$ size(scale) then
      noteIdx ← noteIdx - (2 $\times transaction)$
    end if
    note ← scale[noteIdx]
    notes.append(note, timeBeat + ( noteDuration $\times$ i ), noteDuration, 100)
    i ← i + 1
  end while
  timeBeat ← timeBeat + nBeats
end while
```

These algorithms use three features per beat, and has two threshold arrays.

First it generates a random note in our scale as a first node for our algorithm to start generating on.

Now, per beat, it gets the duration the note is going to have associated with that beat, to do that it sees in which values of our first threshold array the feature is and then returns the duration of the note to be generated.

It also gets the transaction value, the distance between our previous note and our next note. To get this value our algorithm uses the next two features. First, we get a transaction value the same way we got the note duration previously, now with different threshold values, these are associated with the new

Algorithm 5.3: getTransaction

```
f $\leftarrow$ beatFeatures[1]
a $\leftarrow$ beatFeatures[2]
transaction $\leftarrow$ 0
rand $\leftarrow$ random()
while f $>$ threshold2[transaction] do
    transaction $\leftarrow$ transaction + 1
end while
if rand $<= 0.382$ then
    transaction $\leftarrow$ transaction + 0
else if rand $<= 0.532$ then
    transaction $\leftarrow$ transaction + 1
else if rand $<= 0.682$ then
    transaction $\leftarrow$ transaction - 1
else if rand $<= 0.774$ then
    transaction $\leftarrow$ transaction + 2
else if rand $<= 0.866$ then
    transaction $\leftarrow$ transaction - 2
else if rand $<= 0.910$ then
    transaction $\leftarrow$ transaction + 3
else
    transaction $\leftarrow$ transaction - 3
end if
if a $<$ 0 then
    transaction $\leftarrow$ transaction $\times$ (-1)
end if
```

Algorithm 5.4: getNoteDuration

```
vel $\leftarrow$ beatFeatures[0]
if vel $<$ threshold1[0] then
    noteDuration $\leftarrow$ 0.25
else if vel $<$ threshold1[1] then
    noteDuration $\leftarrow$ 0.5
else if vel $<$ threshold1[2] then
    noteDuration $\leftarrow$ 1
else if vel $<$ threshold1[3] then
    noteDuration $\leftarrow$ 2
else
    noteDuration $\leftarrow$ 4
end if
```

feature it is analysing. Then we generate a random number and use a normal distribution to variate the transaction a little so we do not get the same music every time we run our program.

To see if we go up or down the scale we use the last feature, if this feature is positive we go up the scale, if negative we go down. Every time we can not go up or down the scale we go the other way around.

The threshold values were generated by us as we tested the algorithm and saw which values could better describe the movement in our opinions.

With this algorithm we could generate a large number of musics using three features as we wanted.

We decided to generate two musics, each with two tracks.

In our first music, the general movement implementation, uses velocity as feature which determines the note duration, fluidity gives us the transaction and acceleration as weather the transaction is positive or negative in our first track. And for the second track we used contraction index as note duration feature determiner, quantity of movement to generate the transaction and its derivative for the positive or negative transaction value.

The second movement, the specific movement implementation, uses the approach used on the first track of the first music on the upper body features to generate the first track and the same approach for the lower body features for the second track.

In the end this module will generate two array of the type [tempo, instruments, notes] these arrays define the music. Tempo is the bpm value and instruments is a two integer array defining the instruments that will play on track 1 and 2. Notes is array is an array of note arrays, this last one is defined as [track, channel, pitch, time, duration, volume]. Track is the track that note appears in. the channel is always 0 that tells midi what instruments may play that note. The rest of the values were previously explained.

5.5 Implementation merger

In this module we wanted to make a genetic algorithm inspired approach to produce a third piece of music based on our first two musics generated on our previous module.

So, firstly we add all of the notes of both musics on a big array. Now we produced a fitness function to choose which of the notes appear on our third music piece.

This fitness function basically eliminates the duplicate notes and choose upon two different notes played at the same beat the one that is closer to the previous note if our fluidity value is under our threshold, otherwise it chooses the note that is further way from the previous one. If there is no previous note, it means our function has to chose the initial note of our generated music. In this case the function chooses at random out of the two initial notes.

After this we finally used our background attraction value in our approach. We use it as a percentage, and at random we choose that percentage of notes and change them to be out of scale by adding a random value at the pitch value.

The idea behind this was that if you are attracted to the background you do not perceive one hundred percent of the movement. So we introduced some random background noise as if what you hear is also not only coming from the movement in itself.

5.6 Composer

This module receives the three array of notes we composed before, the musical instruments chosen and the bpm value calculated.

With this it uses midiutil.midiFile python package and add the notes to its specific track as he goes through them in our array of notes.

Then as all the notes are now in the midiFile object, it produces our *.midi files as the output of our system.

5.7 Other Implementation Attempts

There were some other features we wanted to add to our system worth mentioning that either did not work as well as we wanted or did not felt like the best way to go as our system was being developed.

As for the fluidity feature, as we said in our section 2, there were to ways to calculate it, we chose the variation of acceleration since when we tried to implement the mass spring damper model approach on its calculation we found that fluidity as a whole body movement was harder to calculate since it only calculate the body members fluidity. And, as everything, our system was some errors when perceiving the body keyjoints so we decided to use the more basic approach to this feature calculation.

When generating our music we also tried to generate a third track with drums as fixed instruments. We wanted to generate a track based on background color analyses. But as implemented and discussed, we decided that drums would push our music to certain music genres. Since our aim for this project is not to develop neither to associate the movement with a certain style nor generate only music of some specific genres. This is a feature that could be pursued in future work around this topic, but not for our project.

A genetic algorithm was thought to be implemented in our project. This was not implemented as we wanted to try out the most associations that made sense to us as we could. Genetic algorithm implementation could be a good implementation to test but it was dropped since our time was limited. This is also a feature that could be implemented and tested in future work.

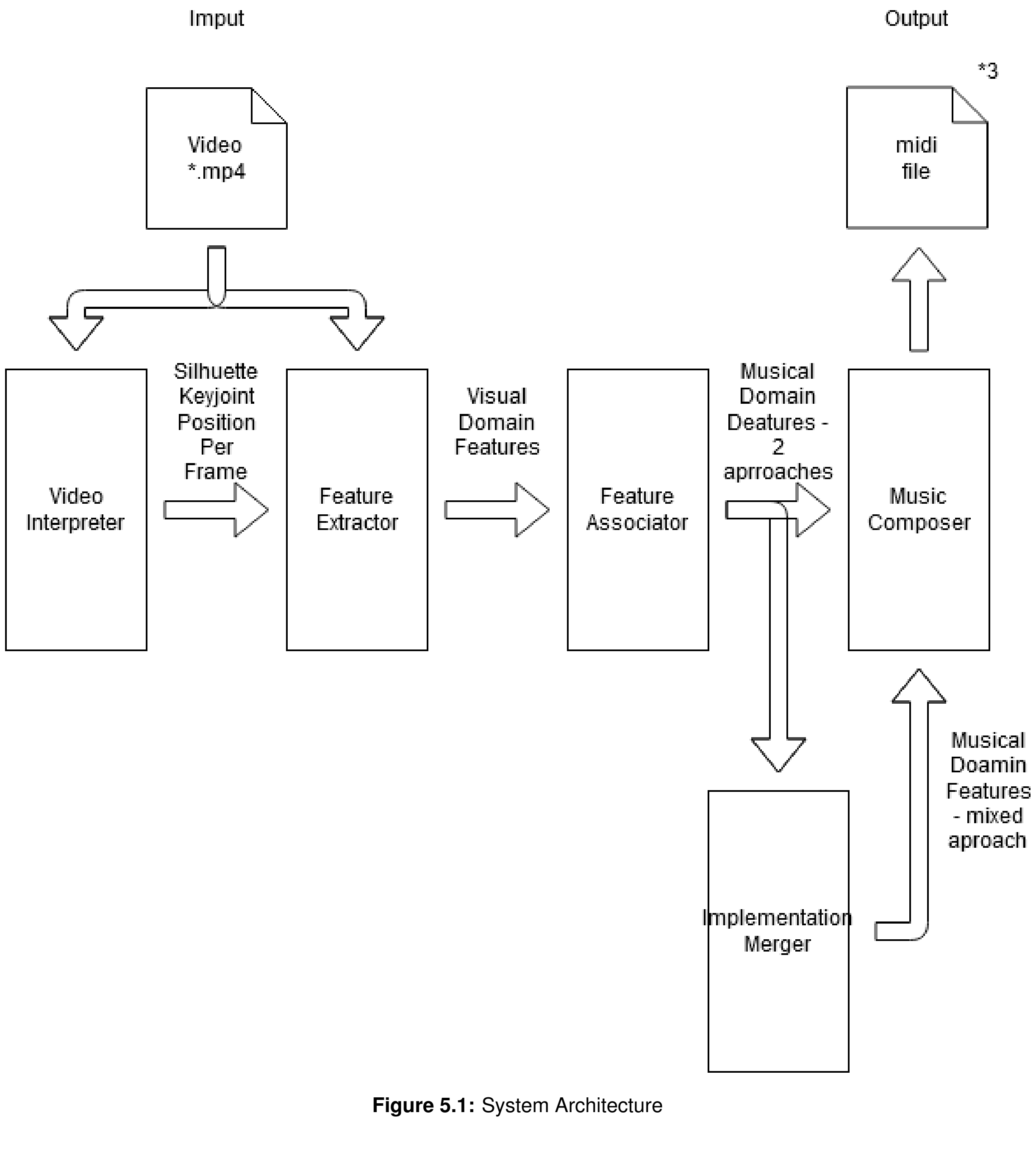


Figure 5.1: System Architecture

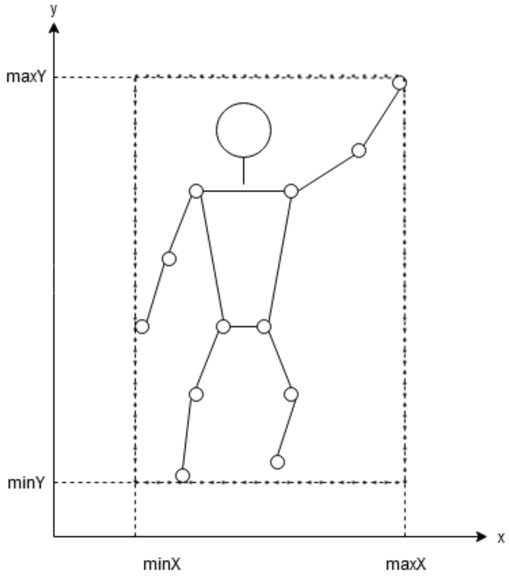


Figure 5.2: bounding box calculation example

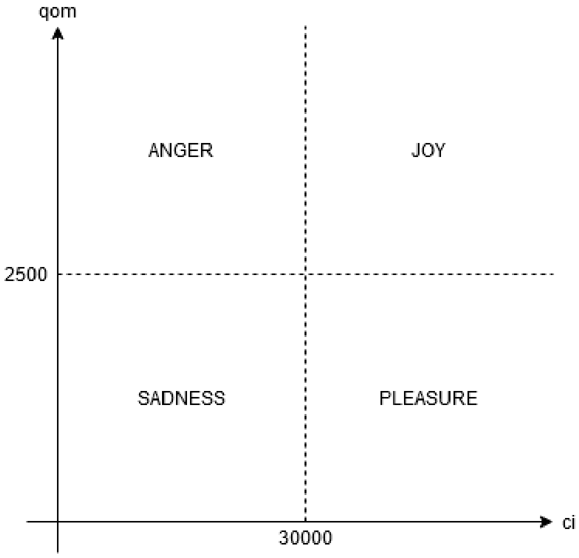


Figure 5.3: Emotion threshold values

6

Case Study

Contents

After our implementation, we need to analyse our results and see if we were able to achieve our goals.

6.1 Data set restrictions

For this project the data set is a set of videos with the following characteristics:

- Non moving single camera
- One person only appearance
- Non moving background
- 1920x1080p resolution
- 30 frames per second

The video restrictions were chosen so that our system could better perceive the movement of the person present in our videos. Non moving single camera helps us with the video processing as well as the non moving background.

Only one person appearance is one of our restrictions as this is a project with limited time frame to be made so we choose to better analyse one person movement instead of multiple person approach. With only one person we could skip the identification of the keyjoint per person which would have to be made in a multiple person approach and also the computations would be more challenging to make.

With this restrictions we were able to find 22 videos with length between 15 seconds and 6 minutes.

With this number we could study different variations of videos.

From different colour videos, different gender and size people and wide and non wide movements.

We chose these variations to be able to better see what our program is able to do and also test its limitations.

6.2 System input & output

As said before, our system receives videos as inputs.

With our restrictions we wanted a variety of videos, found on YouTube, with different features that could give us an overall view about what our system could do in different situations:

- **Ballerina 1 & 2** These videos have 10 seconds each, featuring a woman dancing classic ballet. The main characteristics of these videos are a slow and structured movements with a dark background on the first and a more reddish background on the second one.

- **Conductor**. 29 second video with a conductor conducting his orchestra with his specific movements. A lot of quick, open and closed movements are seen here. In this video we only see the conductor, the background is pitch black.

- **Large man 1 & 2** These are 64 and 78 seconds videos featuring a large sized man with a quick and structured movement in a white background.

- **Woman dancing 1 & 2** Here we see one woman dancing inside a house and on the street so we can evaluate different backgrounds analyses. Their movements are not structure nor choreographed. These videos have between 20 and 40 seconds.

- **Freestyle** This is one of the longest videos processed by our system. A man is the main character, he is doing slow and minor movements with a view of the ocean and horizon contrast.

- **Modern woman** We manage to find a group of videos of a woman dancing a more modern style choreographies in a almost white choreography. Her movements are also structured and variate from slow movements to brutal and quick movements. These videos go from 30 seconds to 6 minutes long

- **Modern man** These videos present the same characteristics as the modern woman ones, but instead of a woman we have a man making similar movements with a darker background.

- **Street dance** Here we see man moving with street dance movements on a multicolored stage.

- **Pumpkin** As our final videos we find a more peculiar video. We have a man dancing with a pumpkin mask on his face. His movements are not structured nor choreographed, he just moves as he pleases. His background is a dark graveyard, has his clothes are also dark colored.

With this group of videos we could analyse multiple different backgrounds and multiple types of movements made by both man and woman of different sizes. Appendix A Show us a frame of each video.

For all these videos, our system produced the three different midi files based on the versions presented in chapter 5.

Since the output is a midi file and one of our goals is to see if we can make music inspired by movement, we had to merge the video and the midi files so that when we could see the video while listening to our generated music.

6.3 System evaluation

To evaluate our system we chose to use questionnaires.

As said our goals are to produce a system capable of produce music based on the human motion present in the video. Our second objective is to see if the music generated match the movement and can be seen as ”inspired by” it. Our final objective has to do with the field we are hoping to contribute to, Computational Creativity. So we aim that this system produce creative objects.

Questionnaires are a good way to see if our system matches our goals, since creativity and music are said to be subjective fields.

In order to not tire our evaluators, we chose to only pick four out of our twenty-two input videos and divide them into two questionnaires. So each questionnaire evaluates two input video implementation outputs at a time, six videos in total (2 videos x 3 implementations = 6 output videos).

Our four choices were based on video diversity, for us to see the evaluation of our system behaviour upon different environments.

So our four choices are:

- **Freestyle**
- **Ballerina-1**
- **Pumpkin**
- **Conductor**

Since our main video feature extraction comes from movement, we chose these four videos since they represent four different and unique types of movements. The first one we see a man moving slowly and with small movements in a field of blue and orange. This video is 3 minutes and 28 seconds long. The second one is a 10 second video that has structured movements characterized by Ballet, these are open and very well choreographed movements. Pumpkin gives us almost the opposite of the ballerina and the Freestyle. This video has not choreographed neither structured movements, just a person dancing as he pleases with big and fast movements. It is 72 seconds long. Finally the Conductor, a 30 second long video, is the only one of our data set that has nothing to do with dance but a lot more to do with music, so we wanted to see what the users think of our systems generated music when reading the movement of the conductor.

So out of our four choices we have small and slow movements, structured movements characterized by a specific dancing style, unstructured big and fast movements and finally a conductor conducting his orchestra.

In our questionnaires, shown on appendix B, we join the first two videos in our first questionnaire and the other two on the second one.

The output videos can be seen in https://www.youtube.com/channel/UCEhuRq75nWCTBhTfPYmIxbA
.

To see if we matched our goals we first ask our evaluators to give us his age, gender and musical knowledge.

Musical knowledge in our questionnaire is divided into four categories:

- None - I only listen to music, don not have any musical education
- Basic - I had some lessons and/or know the basics
- Intermediate - I am attending a music course.
- Advanced - I have a degree in music and/or I am a professional musician

Then they see the three output videos of the first video on that questionnaire and are asked to say which version is their favourite. The evaluator is then asked to select as many adjectives from our adjective list, appendix B as they want to characterize the movement. For the next questions the evaluator is asked to answer with only their favourite approach in mind, so it is asked for the user to describe the sound with their own words and later to select upon the adjective list given previously the ones who better characterize it. Is is also asked on a scale from one to five, being one ”hated it” and five ”loved it”, how much they liked what they heard. Finally in ”yes” or ”no” questions we ask if what the evaluator heard is music and later creative in two different questions.

After this the evaluator listen to the final three videos of that questionnaire and answer to the same questions asked for the previous videos.

This is the structure of our questionnaires that aim to evaluate our system.

6.4 Results

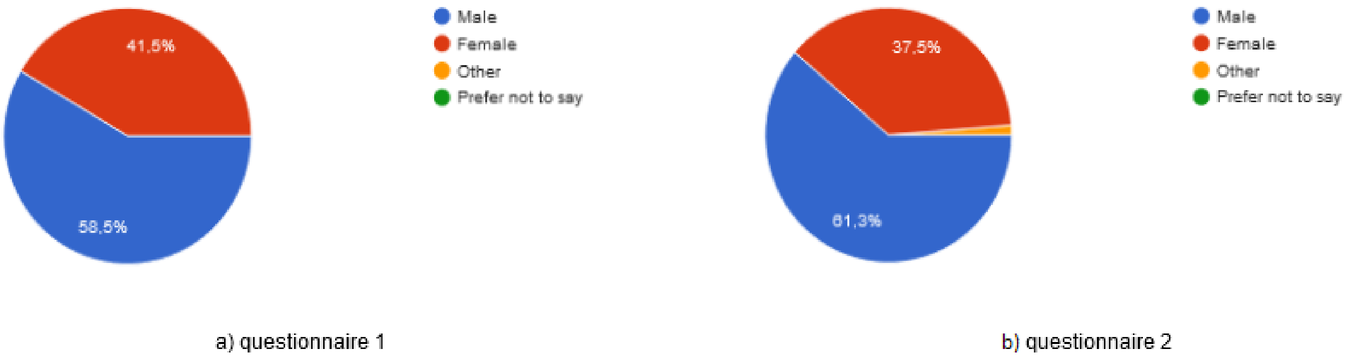
In our first questionnaire featuring Freestyle and Ballerina-1 videos we collected 65 answers. Our second questionnaire was more popular with 80 answers. This brings a total of 155 answers to our questionnaires.

With the answer sample we obtained there are some interesting information that can be extracted.

On the first question we started to ask the evaluators gender. We had a majority of man responding to our forms. Figure 6.1.

Age-wise he had a good diversity with at least one person from every age range answering each form. The majority of our answers came from people with ages between 18-25 and 26-35 years old. Figure 6.2.

Now regarding the musical knowledge we could also get answers from people belonging to all of our four knowledge categories, and as expected most of the people answering had none or basic music knowledge. Figure 6.3.



a) questionnaire 1

b) questionnaire 2

Figure 6.1: Which gender do you most identify yourself?

Table 6.1: Which version did you prefer?

Video	Mean	Median	Mode	Standard deviation
Freestyle	1.82	2	1	0.61
Ballerina-1	2.02	2	2 & 3	0.65
Pumpkin	1.89	2	2	0.57
Conductor	2.24	2.5	3	0.71

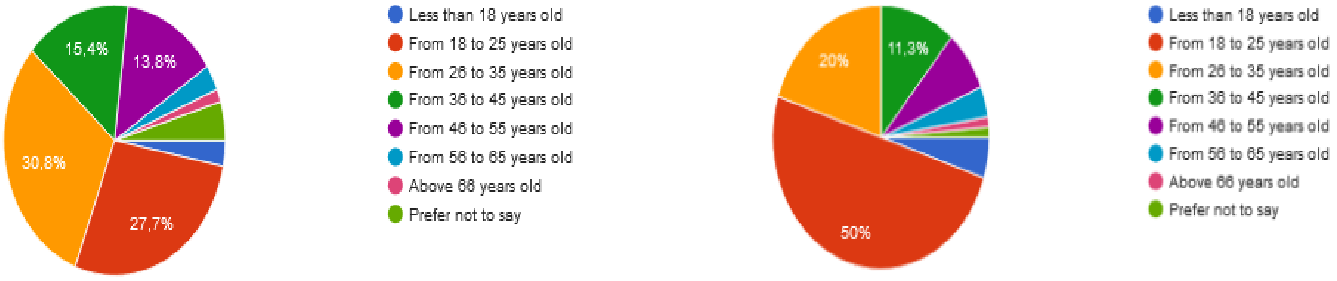
Regarding the preferred version the mode is usually the most relevant measure. But in our case, our mode is either version 1, 2 or 3, and our mean value is always near 2. So we can say that neither of our versions is the preferred one. Table 6.1

When asked if our generated sound is music over 70% of our answers were positive considering what they heard to be music. The conductor video outputs had the most positive answers with 85% of the answers agreeing with what they heard can be considered music. Figure 6.4

Now that we established that most of our answers say that the sound is music, in table 6.2 we see the mean, median, mode and standard deviation values of how much our evaluators liked our generated music. Looking at the mean and median the values are all around the number 3 which is a neutral, the users neither hated nor loved our music. The mode values variate between 3 and 4 which tells us that there are a lot of neutral and positive answers but since our average is around or bellow three it means that there are some people that put the number 1, hated the music this means.

Now looking at the question if weather or not the users found the music to be creative we can see more or less the same answers given to weather the user found the sound to be considered music. Over 70% of our answers were positive, which means we can consider our system to be considered as creative. Figure 6.5

As said before the evaluator is asked to describe both the movement and sound of their favourite



a) questionnaire 1

b) questionnaire 2

Figure 6.2: What is your age?

Table 6.2: How much did you like the sound? (being 1 hated it and 5 loved it)

Video	Mean	Median	Mode	Standard deviation
Freestyle	2.86	3	3	1.01
Ballerina-1	3.05	3	4	1.05
Pumpkin	2.74	3	4	1.24
Conductor	2.96	3	3	0.91

version from a list of adjectives given by us. From these questions we can see if the sound generated is related with the movement seen.

For the first video of the first questionnaire, we have 129 common adjectives between movement and sound user answers, this represent 66% of the total distinct answers given to both movement and sound description. Figure 6.6.

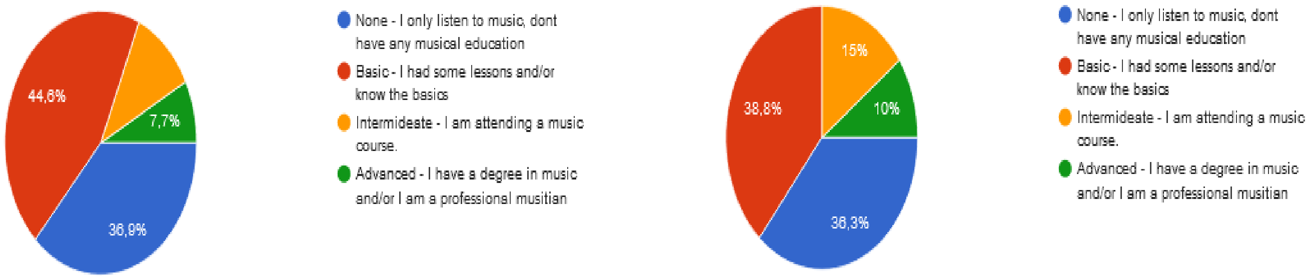
The ballerina-1 video is the one with the most different distinct adjectives answers, yet more than half, 56% of the adjectives given are common in both movement and sound description. Figure 6.7.

For both of the videos present on questionnaire 2 we have 66% of common adjectives on both descriptions. Figures 6.8 and 6.9.

With these results we can say that over 60% of our sound adjectives match the adjectives given to the movement. So it is safe to say that our sound can be considered related with the movement it was generated on.

About the open question that we ask to describe with their own words the sound they hear we found these answers to be interesting for our evaluation:

- ”It’s a calm sound, with some higher notes but slightly inharmonious. Repetitive but not having a clear melody.” regarding freestyle video.



a) questionnaire 1

b) questionnaire 2

Figure 6.3: How would you describe your musical knowledge?

- ”this brought back to my memory the sounds coming out from a Nokia 3310, it is not the kind of music i would listen to on a vacay road trip, but it is indeed a creative piece.” regarding freestyle video.

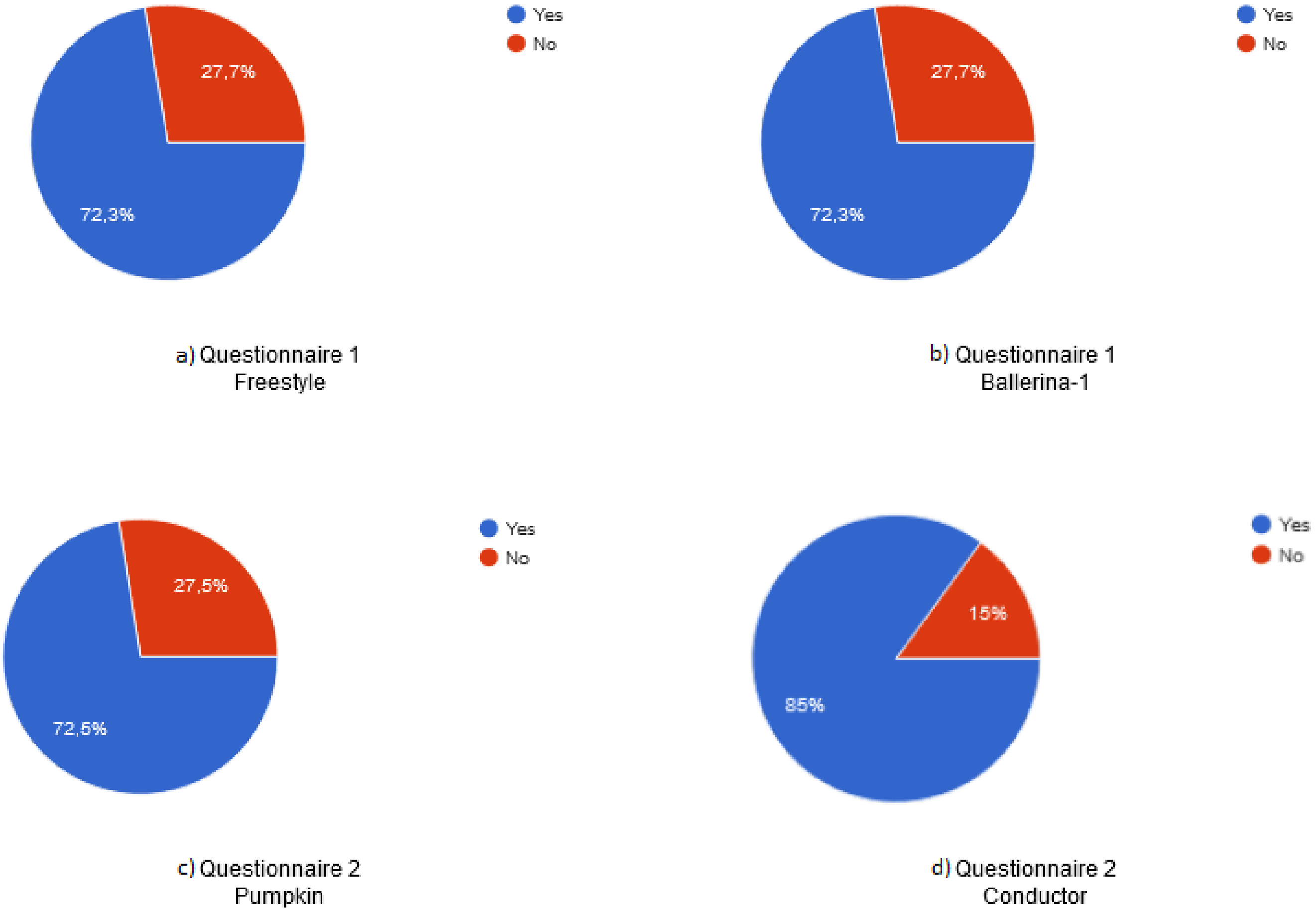
- ”Rather inscrutable and not particularly tuneful, but structured and smooth-sounding. It was like trying to listen to a conversation in a language that you don’t understand almost at all – interesting at first, but a bit frustrating after a while when you fail to get some kind of meaning out of it.” regarding freestyle video

- ”To me, the music sounds rather difficult to understand or get any emotion from. Maybe it makes sense to a music theorist who can pick out all the clever note combinations and changes in key and such, but to me it feels empty, as with the previous music. There is a vaguely cheerful quality to it, but it’s artificial.” regarding ballerina-1 video

- ”It was a bit confusing at first to listen to the song, but then suddenly it becomes interesting and enjoyable both movement and sound.” regarding pumpkin video.

- ”The tone of the instruments made the music be sort of melancholic and sad, although the melody made it be energetic. The combination of those two characteristics gave me the impression of it being composed by a sad person who is spilling out all of his energy on this song.” regarding pumpkin dance

- ”It sounds like AI made music, it’s honestly how I would describe this piece of audio if I heard it out of context. Or as a piece from a very inexperienced composer who just downloaded Sibelius. It’s definitely does not sound like a random array of sounds, but it’s like the calculation of the



a) Questionnaire 1
Freestyle

b) Questionnaire 1
Ballerina-1

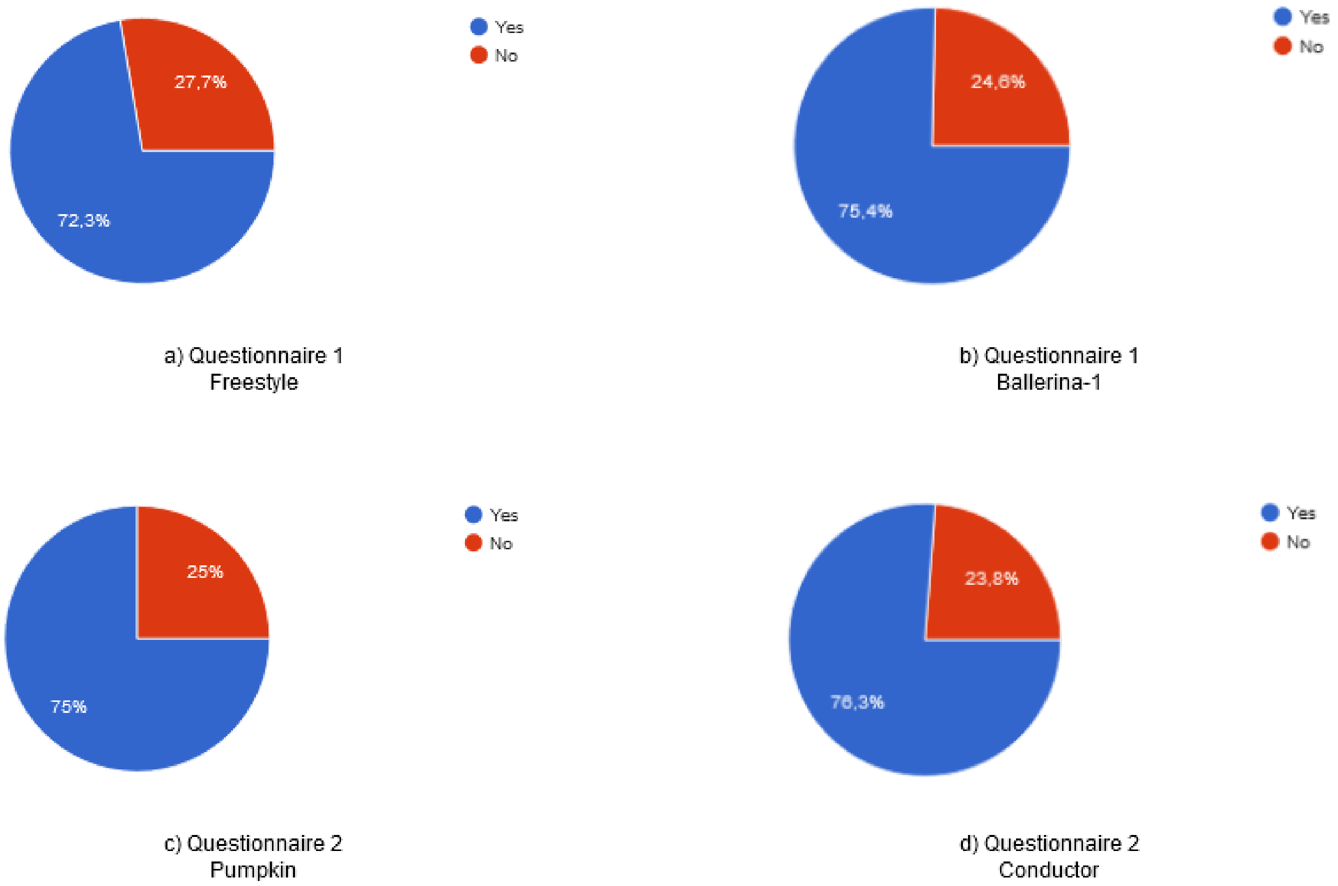
c) Questionnaire 2
Pumpkin

d) Questionnaire 2
Conductor

Figure 6.4: Do you consider what you heard to be music?

successive intervals is random in a way, because there are too many leaps following each other for it to sound coherent. I feel as if there is not enough repetition or recycling of melodic material but a constant almost random generation of new material. This makes the piece one blob of indistinguishable character. There is no sense of chordal harmony in the overlap of melodies, and it makes it sound all the same. The melodic rhythm however changes somewhat with the movement, but the difference in gestures almost suggests a formal novelty like it asked for new musical part or repetition, and what we hear is a mere quickening of melodic rhythm, it's as if there is a divorce between sight and ear that causes a sense of strangeness in the hole thing." regarding pumpkin video.

- "The fact that there is a drone note in a middle voice gives the whole thing a sense of drive and direction and when it vanishes it's absence causes a sense of no ground, like the music was in a mine cart and the rails where left broken in the air and the cart just drove ahead falling somewhere" regarding conductor video



a) Questionnaire 1
Freestyle

b) Questionnaire 1
Ballerina-1

c) Questionnaire 2
Pumpkin

d) Questionnaire 2
Conductor

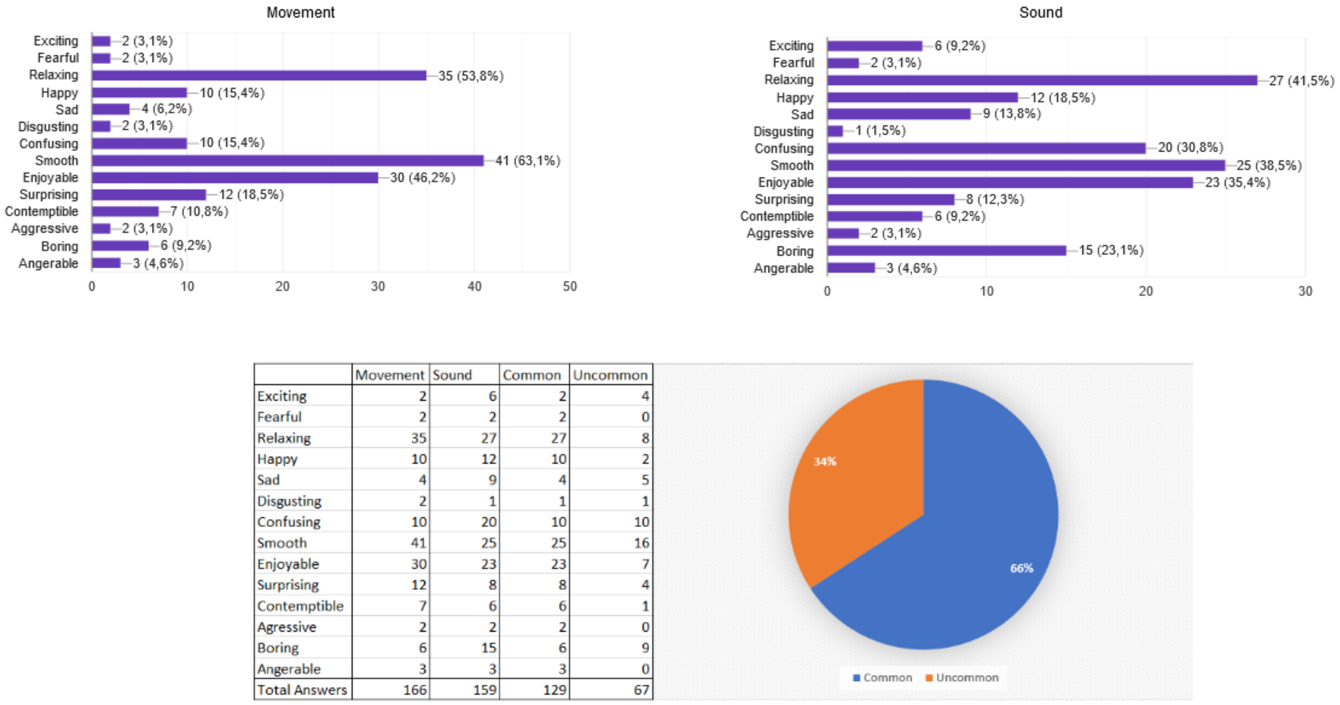
Figure 6.5: Do you consider what you heard to be creative?

Here we see that the evaluators find the music pieces to be creative but there is something missing on it. Some of them suggest that music structure is missing and that is what causes the ”weird” feeling. With this question we can see a more clear opinion of our evaluators with some ideas of how to proceed in order to make this project more appealing to the general audience.

when filtering the results by gender, age range and musical knowledge we could not find interesting or significant results to make new assumptions. This means that our question do not have a age, gender or musical knowledge bias.

Analysing the results, over 70% of our answers say that our output sound is either music and a creative object when considering their favourite version. Since we do not have a preferred version we could say that so we can say that for all our implementations both of our first two goals were achieved. Our system is able to generate pieces that are considered music and creative pieces for the majority of people that evaluated our system.

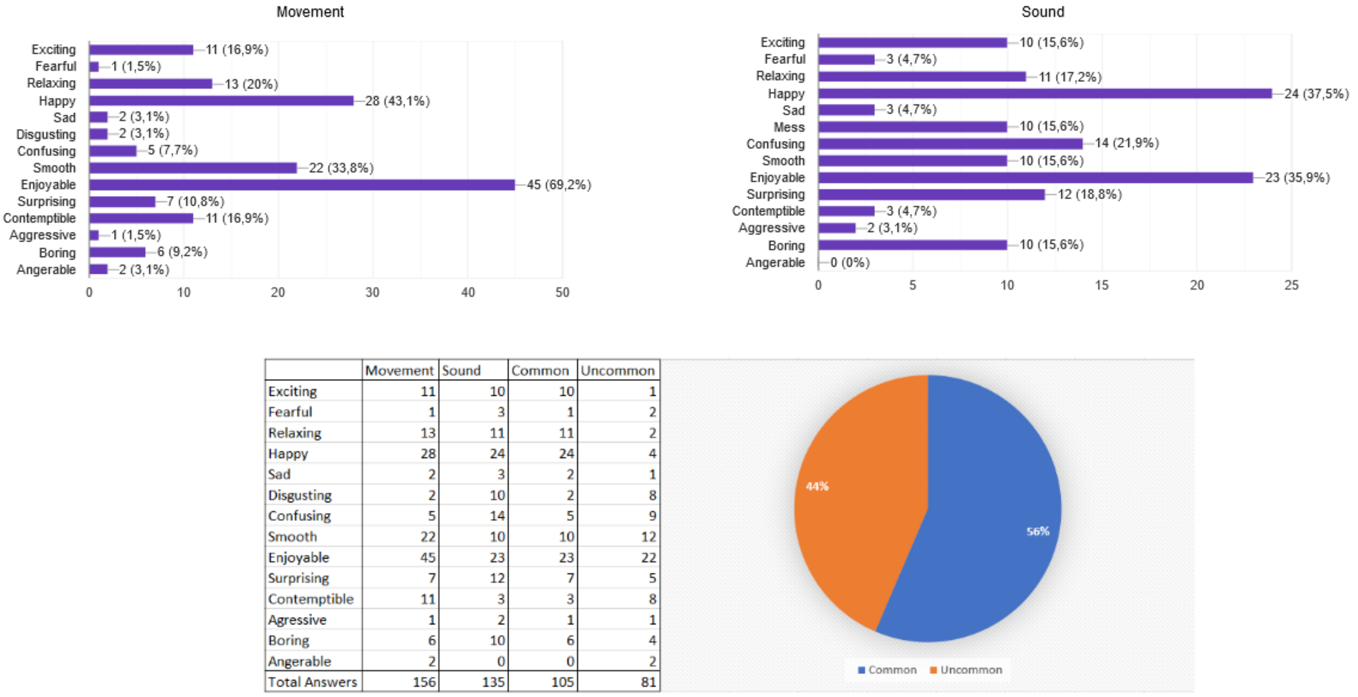
For the movement to sound association goal we can prove that our users found the sound to be related to the movement perceived. So this final goal can also be said to be achieved.



	Movement	Sound	Common	Uncommon
Exciting	2	6	2	4
Fearful	2	2	2	0
Relaxing	35	27	27	8
Happy	10	12	10	2
Sad	4	9	4	5
Disgusting	2	1	1	1
Confusing	10	20	10	10
Smooth	41	25	25	16
Enjoyable	30	23	23	7
Surprising	12	8	8	4
Contemptible	7	6	6	1
Agressive	2	2	2	0
Boring	6	15	6	9
Angerable	3	3	3	0
Total Answers	166	159	129	67

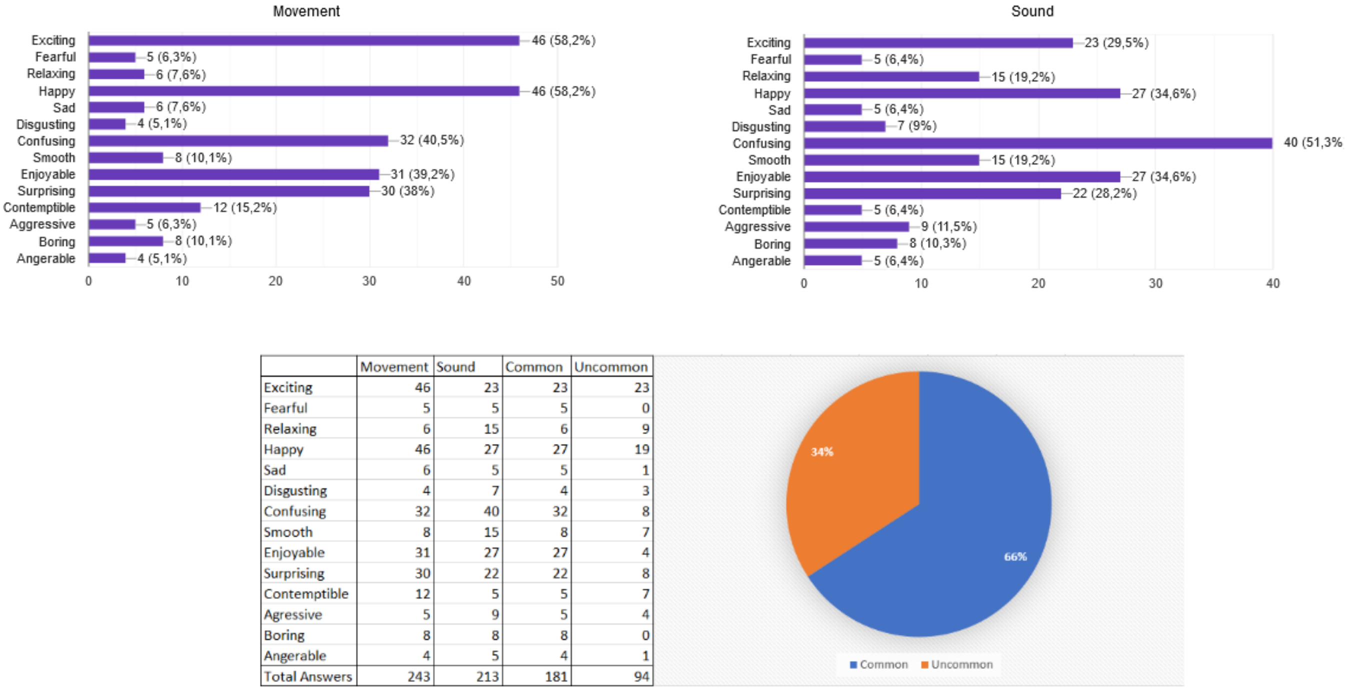
Figure 6.6: freestyle movement-sound association

Since we tested our system ability to perceive and generate music in four different environments, and there were simple variations on our results, we could say that overall, our system is capable to have satisfactory results upon the tested environments.



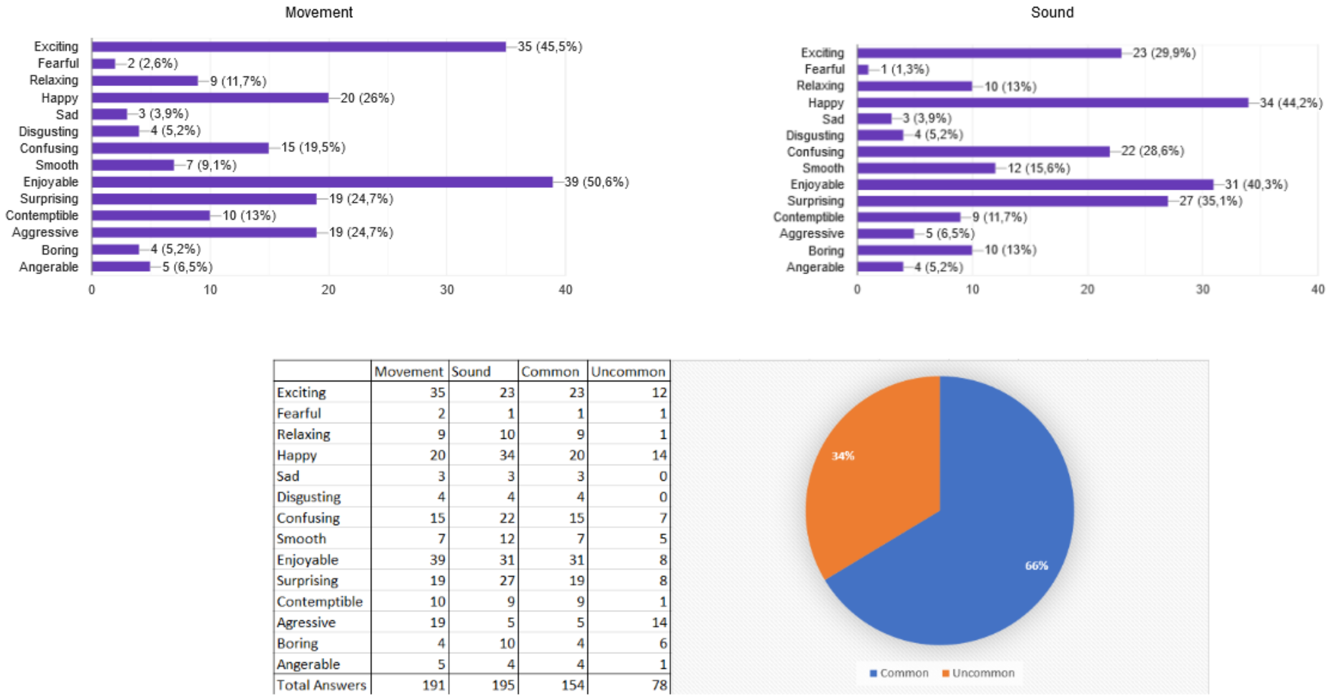
	Movement	Sound	Common	Uncommon
Exciting	11	10	10	1
Fearful	1	3	1	2
Relaxing	13	11	11	2
Happy	28	24	24	4
Sad	2	3	2	1
Disgusting	2	10	2	8
Confusing	5	14	5	9
Smooth	22	10	10	12
Enjoyable	45	23	23	22
Surprising	7	12	7	5
Contemptible	11	3	3	8
Agressive	1	2	1	1
Boring	6	10	6	4
Angerable	2	0	0	2
Total Answers	156	135	105	81

Figure 6.7: ballerina-1 movement-sound association



	Movement	Sound	Common	Uncommon
Exciting	46	23	23	23
Fearful	5	5	5	0
Relaxing	6	15	6	9
Happy	46	27	27	19
Sad	6	5	5	1
Disgusting	4	7	4	3
Confusing	32	40	32	8
Smooth	8	15	8	7
Enjoyable	31	27	27	4
Surprising	30	22	22	8
Contemptible	12	5	5	7
Agressive	5	9	5	4
Boring	8	8	8	0
Angerable	4	5	4	1
Total Answers	243	213	181	94

Figure 6.8: pumpkin movement-sound association



	Movement	Sound	Common	Uncommon
Exciting	35	23	23	12
Fearful	2	1	1	1
Relaxing	9	10	9	1
Happy	20	34	20	14
Sad	3	3	3	0
Disgusting	4	4	4	0
Confusing	15	22	15	7
Smooth	7	12	7	5
Enjoyable	39	31	31	8
Surprising	19	27	19	8
Contemptible	10	9	9	1
Agressive	19	5	5	14
Boring	4	10	4	6
Angerable	5	4	4	1
Total Answers	191	195	154	78

Figure 6.9: Conductor movement-sound association

7

Conclusions and Future Work

There are a number of different approaches one can take to try to understand creativity [35]. It is an aspect of human intelligence that has a role in our problem solving ability and so in our day to day life [5].

The simple evaluation of weather a piece is creative or not can be a simple task, but the justification behind weather that piece is or not creative is not. To explain why something is creative we have to be able to first define what creativity is which is a much more challenging task than the first one.

This thesis proposed, developed and tested a system that perceives human body motion present in videos and generates music based on that perception. Our main goals are to create sound pieces that could be seen as movement inspired, for those sound pieces to be considered music inspired by the movement and finally for that music to be considered a creative piece in order to our system to be considered creative as well.

First we learned what visual features are present in a video and human body motion, and out of those features how can we capture them and use them as ”inspiration” for our music generation.

Then we needed to gain some musical knowledge to be able to produce something to be later evaluated as music or not. Musical theory played a huge role in our thesis development.

The obtained knowledge is explained in this document as well as in previous work we analysed that gave us some ideas about what to do and what choices to make as when developing our system.

Since our aim is to develop a system capable of perceiving human body motion and with that generate music, our system input is so a video with fixed camera and only one person appearance. Our system then processes the data and generate midi files containing our generated music.

Regarding our implementation we use different techniques of computer vision, artificial intelligence, mathematics to be able to develop our system. In the end we were able to produce three music midi files each with a different movement to sound association.

One association sees the whole movement as one and generates music based on that assumption. The second one sees the human movement from the waist up and from the waist down as two different movements and compose music based on that. The final one assumes that both of our previous associations have a point but the background has some effect in what we perception. So basically it merges the two previous associations into one and lithely changes the merged association based on the background analyses.

For the evaluation of our work, we used questionnaires as a way to evaluate it based on our goals.

Based on 155 answers we could say that over than 70% of our answers considered our output pieces to be considered music and creative. Over 60% of the answers given could lead us to say that overall our music is characterized as related with the movement our evaluators perceive. We can say that our goals were reached with the development of this system.

As said in the title, our system is a computational creativity approach to the movement to sound association. There can be an almost infinite different ways to make a system capable of what our system is capable of.

Our system has its weak points, some users found the music to be too repetitive, without feeling transmission or without a structure. As future work we propose these aspects to be considered.

Our system is capable of capturing the movement of one person, yet the software used for human body capturing is capable of multiple person capturing. A system capable of perceiving multiple movements and mix them in a creative way is also a way to further develop our system.

Machine learning is also used for music structure. As this is something our evaluators mentioned or system lacked. An approach of structuring our final pieces using machine learning could be an interesting approach to be made.

Bibliography

[1] Achyuta Aich, Tanwi Mallick, Himadri BGS Bhuyan, Partha Pratim Das, and Arun Kumar Majumdar. Nrityaguru: A dance tutoring system for bharatanatyam using kinect. In *Proceedings. National Conference on Computer Vision, Pattern Recognition, Image Processing, and Graphics*, pages 481–493, 2017.

[2] Kenji Amaya, Armin Bruderlin, and Tom Calvert. Emotion from motion. In *Proceedings. Graphics interface*, pages 222–229, 1996.

[3] Matthew S Biagini, Lee E Brown, Jared W Coburn, Daniel A Judelson, Traci A Statler, Martim Bottaro, Tai T Tran, and Nick A Longo. Effects of self-selected music on strength, explosiveness, and mood. *The Journal of Strength & Conditioning Research*, 26:1934–1938, 2012.

[4] Kim Binsted, Helen Pain, and Graeme D Ritchie. Children's evaluation of computer-generated punning riddles. *Pragmatics & Cognition*, 5(2):305–354, 1997.

[5] Margaret A Boden. Computer models of creativity. *AI Magazine*, 30(3):23–34, 2009.

[6] Georg Boenn, Martin Brain, Marina De Vos, et al. Automatic composition of melodic and harmonic music by answer set programming. In *Proceedings. International Conference on Logic Programming*, pages 160–174, 2008.

[7] Ginevra Castellano, Santiago D Villalba, and Antonio Camurri. Recognising human emotions from body movement and gesture dynamics. In *Proceedings. International Conference on Affective Computing and Intelligent Interaction*, pages 71–82, 2007.

[8] Keunwoo Choi, George Fazekas, and Mark Sandler. Text-based lstm networks for automatic music composition. *arXiv preprint arXiv:1604.05358*, 2016.

[9] Florian Colombo, Samuel P Muscinelli, Alexander Seeholzer, Johanni Brea, and Wulfram Gerstner. Algorithmic composition of melodies with deep recurrent neural networks. *arXiv preprint arXiv:1606.07251*, 2016.

[10] Georges Dalleau, Alain Belli, Muriel Bourdin, and Jean-René Lacour. The spring-mass model and the energy cost of treadmill running. *European journal of applied physiology and occupational physiology*, 77(3):257–263, 1998.

[11] Douglas Eck and Juergen Schmidhuber. Finding temporal structure in music: Blues improvisation with lstm recurrent networks. In *Proceedings of the 12th IEEE workshop on neural networks for signal processing*, pages 747–756, 2002.

[12] Claudio Fanti, Lihi Zelnik-Manor, and Pietro Perona. Hybrid models for human motion recognition. *Proceedings. IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR 200*, 1:1166–1173, 2005.

[13] Jose D Fernández and Francisco Vico. Ai methods in algorithmic composition: A comprehensive survey. *Journal of Artificial Intelligence Research*, 48:513–582, 2013.

[14] Emma Frid, Ludvig Elblaus, and Roberto Bresin. Interactive sonification of a fluid dance movement: an exploratory study. *Journal on Multimodal User Interfaces*, 13(3):181–189, 2019.

[15] Dariu M Gavrila and Larry S Davis. 3-d model-based tracking of humans in action: a multi-view approach. In *Proceedings cvpr ieee computer society conference on computer vision and pattern recognition*, pages 73–80, 1996.

[16] Olivier Girard, Jean-Paul Micallef, and Grégoire P Millet. Changes in spring-mass model characteristics during repeated running sprints. *European journal of applied physiology*, 111(1):125–134, 2011.

[17] E Mark Gold. Language identification in the limit. *Information and control*, 10(5):447–474, 1967.

[18] Rafael Gonzalez and Richard Woods. Prentice-Hall, Inc., 2006.

[19] Edward H Hagen and Gregory A Bryant. Music and dance as a coalition signaling system. *Human nature*, 14:21–51, 2003.

[20] Hermann Hild, Johannes Feulner, and Wolfram Menzel. Harmonet: A neural net for harmonizing chorales in the style of js bach. In *Proceedings. Advances in neural information processing systems*, pages 267–274, 1992.

[21] Lejaren A Hiller Jr and Leonard M Isaacson. Musical composition with a high-speed digital computer. *Journal of the Audio Engineering Society*, 6(3):154–160, 1958.

[22] Sevinc Kurt and Kelechi Kingsley Osueke. The effects of color on the moods of college students. *SAGE Open*, 4(1):2158244014525423, 2014.

[23] Elad Liebman, Peter Stone, and Corey N White. How music alters decision making-impact of music stimuli on emotional classification. In *ISMIR*, pages 793–799, 2015.

[24] Tanwi Mallick, Partha Pratim Das, and Arun Kumar Majumdar. Posture and sequence recognition for bharatanatyam dance performances using machine learning approach. *arXiv preprint arXiv:1909.11023*, 2019.

[25] Michael C Mozer. Induction of multiscale temporal structure. In *Proceedings. Advances in neural information processing systems*, pages 275–282, 1992.

[26] Carolyn J Murrock and Patricia A Higgins. The theory of music, mood and movement to improve health outcomes. *Journal of Advanced Nursing*, 65:2249–2257, 2009.

[27] Luciana Porcher Nedel and Daniel Thalmann. Real time muscle deformations using mass-spring systems. In *Proceedings. Computer Graphics International (Cat. No. 98EX149)*, pages 156–165, 1998.

[28] Joseph O'rourke and Norman I Badler. Model-based image analysis of human motion using constraint propagation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2(6):522–536, 1980.

[29] Stefano Piana, Paolo Alborno, Radoslaw Niewiadomski, Maurizio Mancini, Gualtiero Volpe, and Antonio Camurri. Movement fluidity analysis based on performance and perception. In *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, pages 1629–1636, 2016.

[30] Gerard J Puccio. From the dawn of humanity to the 21st century: creativity as an enduring survival skill. *The Journal of Creative Behavior*, 51:330–334, 2017.

[31] R Keith Sawyer. *Explaining creativity: The Science of Human Innovation*. Oxford University Press, 2011.

[32] Ahmad S Shaarani and Daniela M Romano. Emotional body movements. *HUMAINE Summer School*, 22, 2006.

[33] Cristian Sminchisescu, Atul Kanaujia, and Dimitris Metaxas. Conditional models for contextual human motion recognition. *Computer Vision and Image Understanding*, 104(2-3):210–220, 2006.

[34] Yang Song, Luis Goncalves, and Pietro Perona. Unsupervised learning of human motion. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 25(7):814–827, 2003.

[35] Robert J Sternberg. The nature of creativity. *Creativity research journal*, 18:87, 2006.

[36] Frederik Styns, Leon van Noorden, Dirk Moelants, and Marc Leman. Walking on music. *Human movement science*, 26(5):769–785, 2007.

[37] Joana Teixeira. Cross-domain analogy from image to music. Master's thesis, Instituto Superior Técnico, Universidade de Lisboa, 2017.

[38] Christine Marie Tipper, Giulia Signorini, and Scott T Grafton. Body language in the brain: constructing meaning from expressive movement. *Frontiers in Human Neuroscience*, 9:450, 2015.

A

Frames of videos

Figure A.1: Ballerina-1

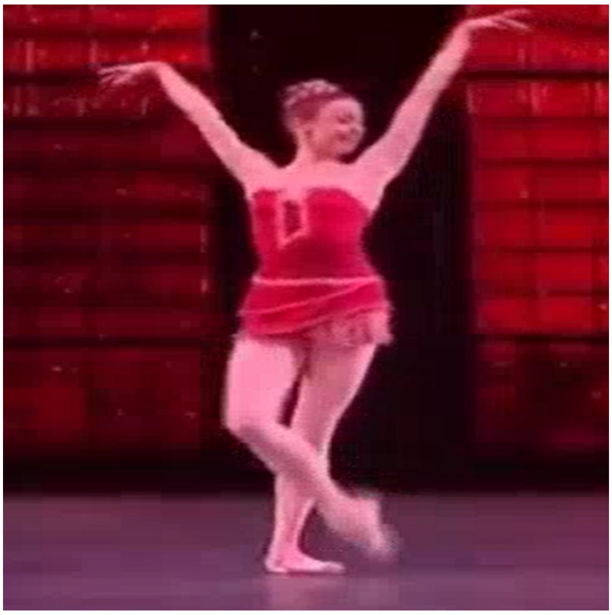


Figure A.2: Ballerina-2

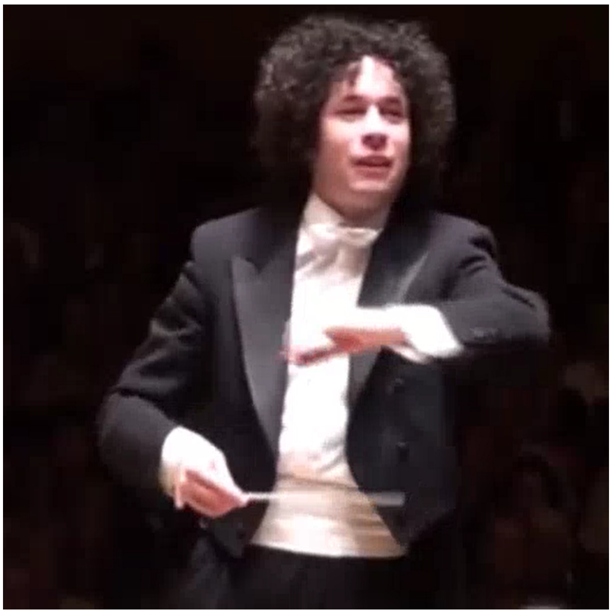


Figure A.3: Conductor

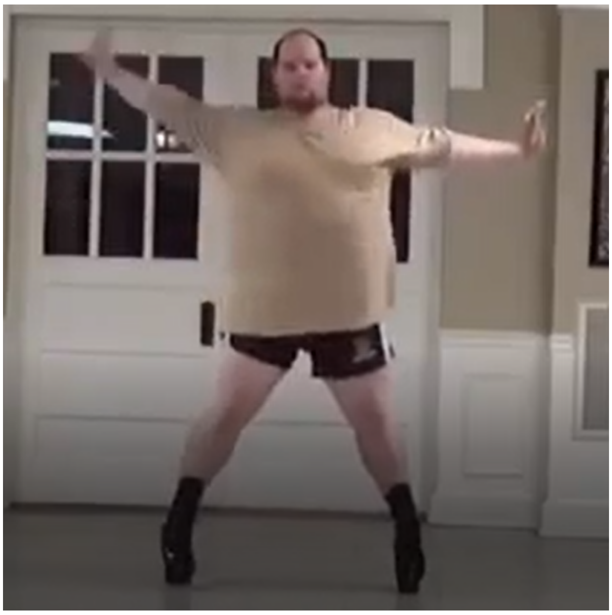


Figure A.4: Large-man-1

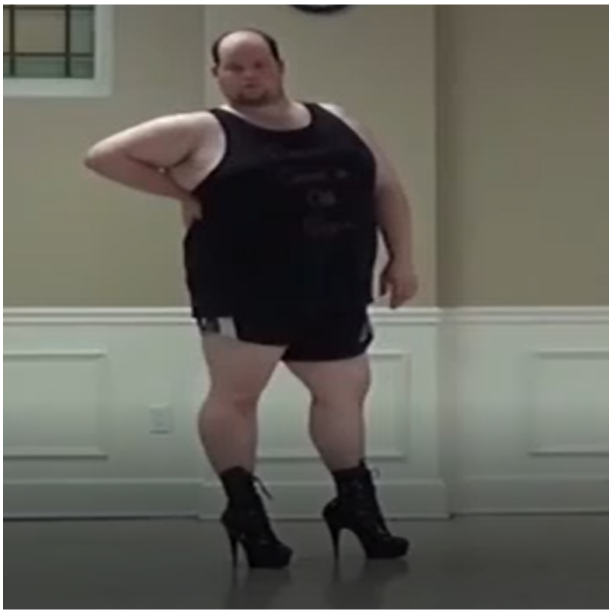


Figure A.5: Large-man-2

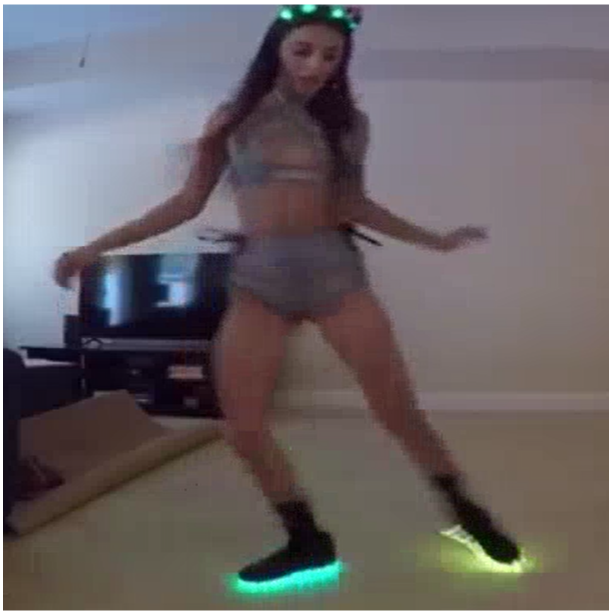


Figure A.6: Woman-dancing-1

Figure A.7: Woman-dancing-2

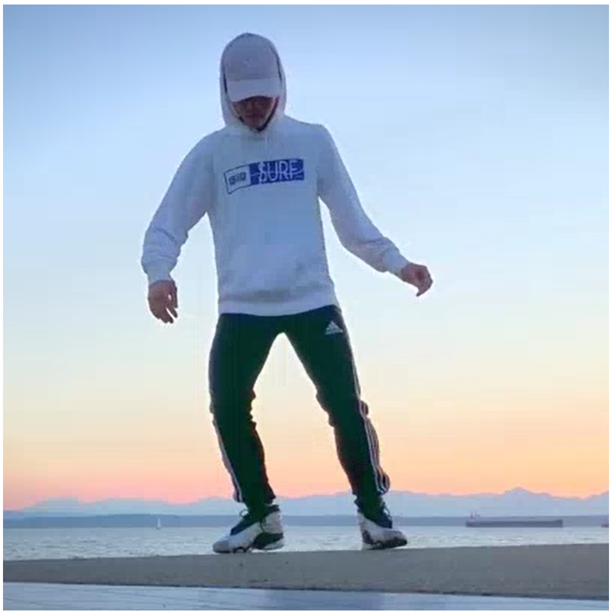


Figure A.8: Freestyle

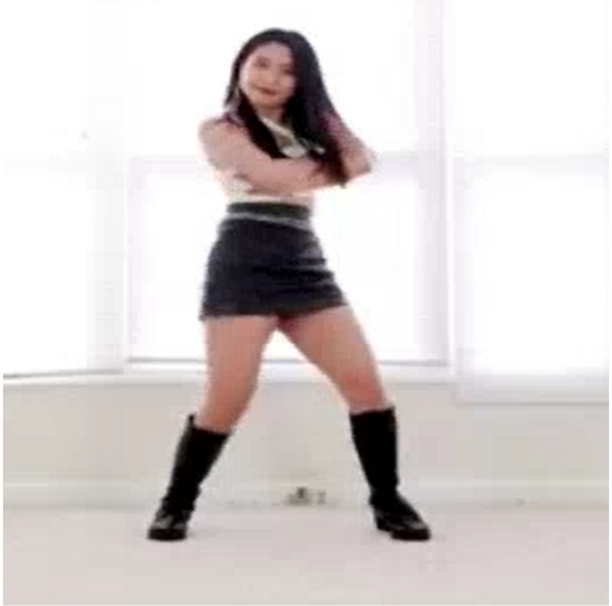


Figure A.9: Modern-woman

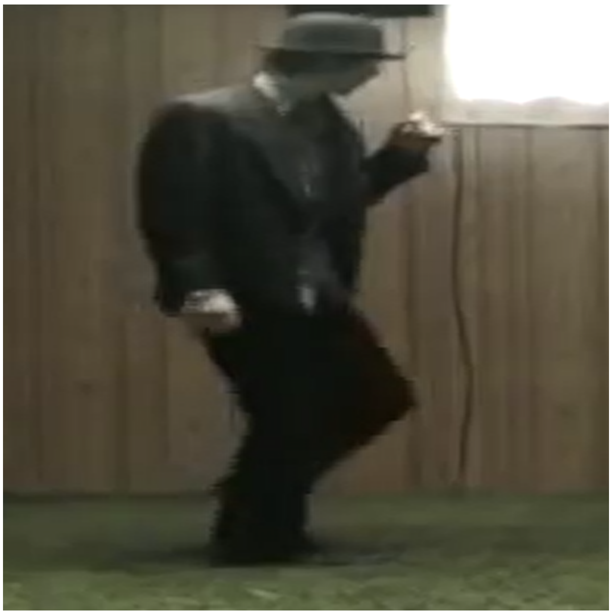


Figure A.10: Modern-man

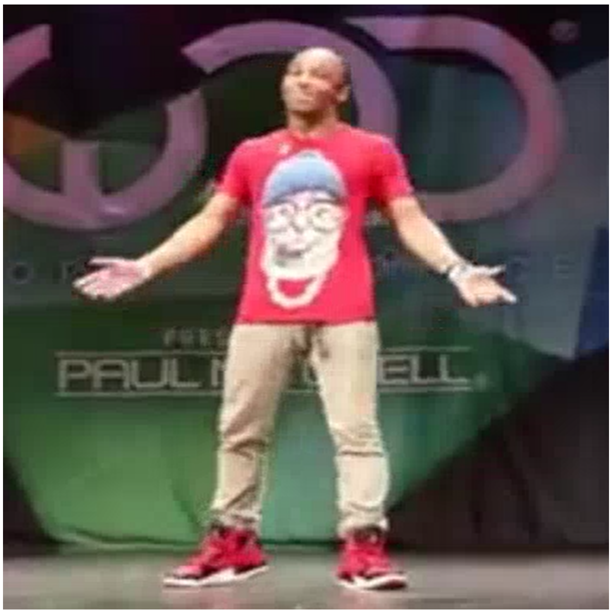


Figure A.11: Street-dance

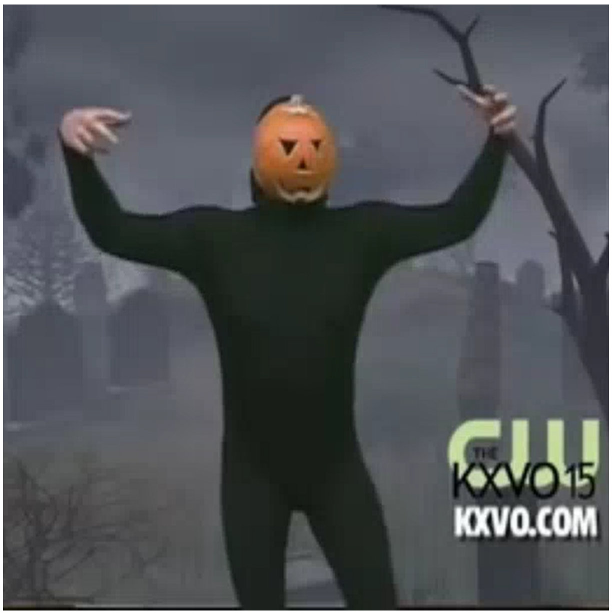
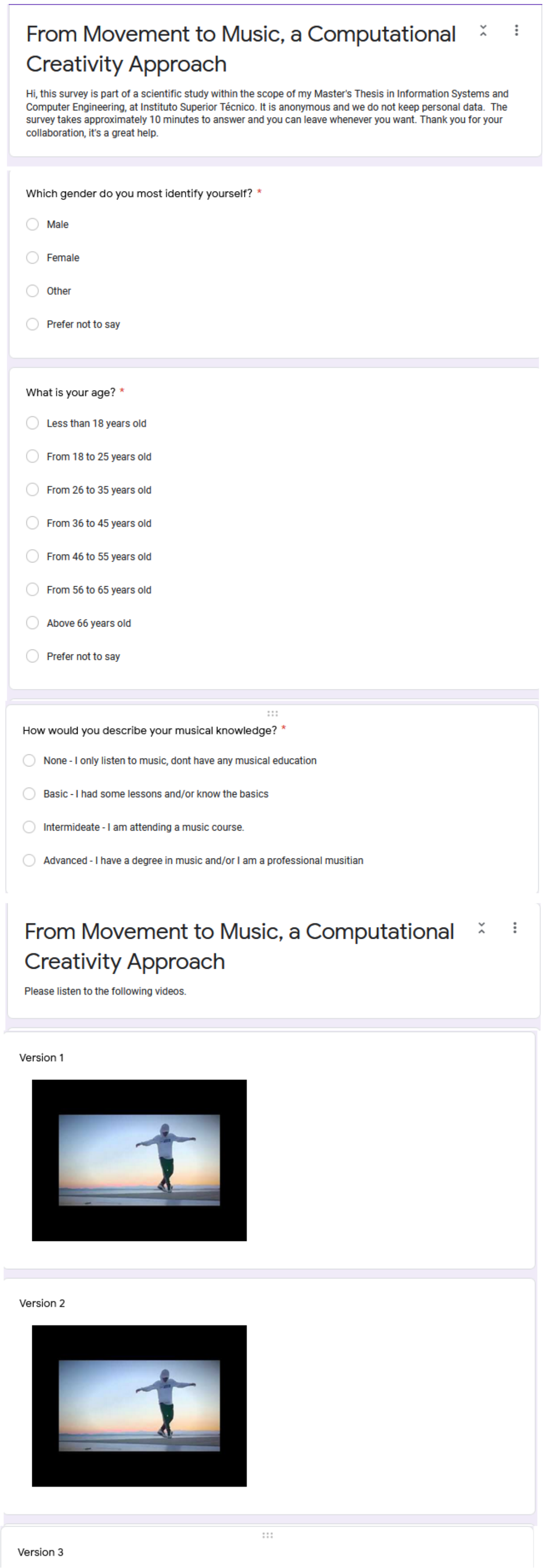


Figure A.12: Pumpkin

B

Questionnaires



From Movement to Music, a Computational Creativity Approach

Hi, this survey is part of a scientific study within the scope of my Master's Thesis in Information Systems and Computer Engineering, at Instituto Superior Técnico. It is anonymous and we do not keep personal data. The survey takes approximately 10 minutes to answer and you can leave whenever you want. Thank you for your collaboration, it's a great help.

Which gender do you most identify yourself? *

○ Male

○ Female

○ Other

○ Prefer not to say

What is your age? *

○ Less than 18 years old

○ From 18 to 25 years old

○ From 26 to 35 years old

○ From 36 to 45 years old

○ From 46 to 55 years old

○ From 56 to 65 years old

○ Above 66 years old

○ Prefer not to say

How would you describe your musical knowledge? *

○ None - I only listen to music, dont have any musical education

○ Basic - I had some lessons and/or know the basics

○ Intermideate - I am attending a music course.

○ Advanced - I have a degree in music and/or I am a professional musitian

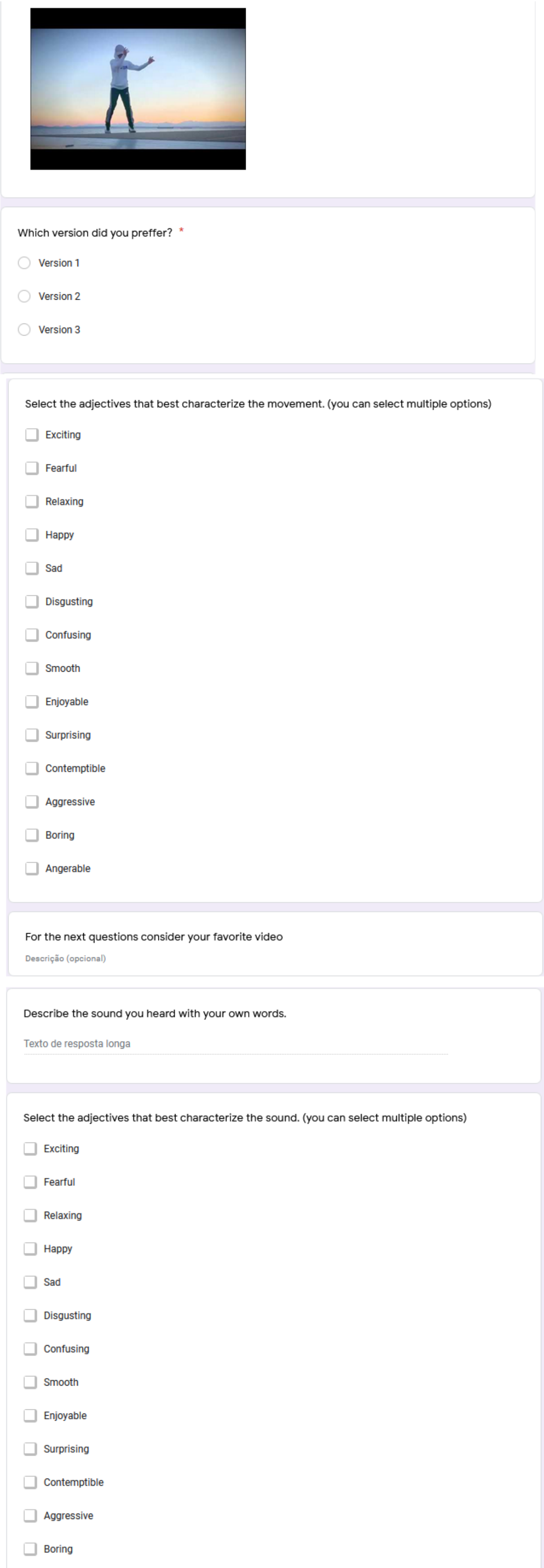
From Movement to Music, a Computational Creativity Approach

Please listen to the following videos.

Version 1

Version 2

Version 3



Which version did you preffer? *

○ Version 1

○ Version 2

○ Version 3

Select the adjectives that best characterize the movement. (you can select multiple options)

□ Exciting

□ Fearful

□ Relaxing

□ Happy

□ Sad

□ Disgusting

□ Confusing

□ Smooth

□ Enjoyable

□ Surprising

□ Contemptible

□ Aggressive

□ Boring

□ Angerable

For the next questions consider your favorite video

Descrição (opcional)

Describe the sound you heard with your own words.

Texto de resposta longa

Select the adjectives that best characterize the sound. (you can select multiple options)

□ Exciting

□ Fearful

□ Relaxing

□ Happy

□ Sad

□ Disgusting

□ Confusing

□ Smooth

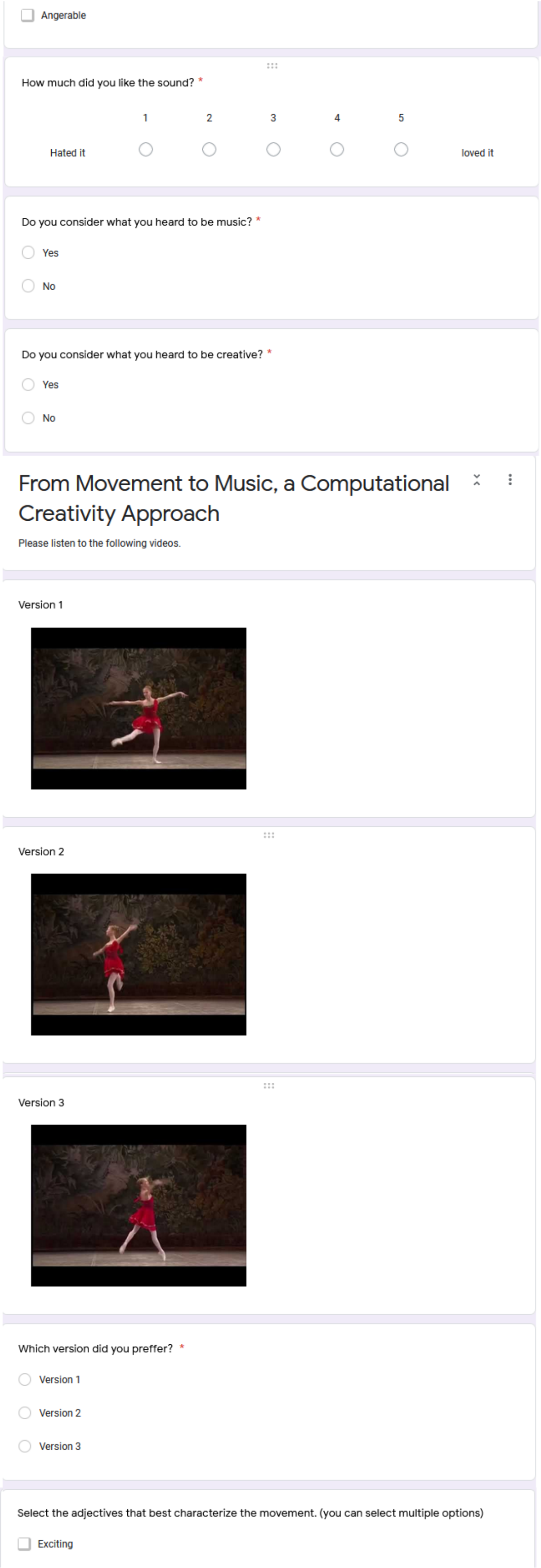
□ Enjoyable

□ Surprising

□ Contemptible

□ Aggressive

□ Boring



□ Angerable

How much did you like the sound? *

	1	2	3	4	5	
Hated it	○	○	○	○	○	loved it

Do you consider what you heard to be music? *

○ Yes

○ No

Do you consider what you heard to be creative? *

○ Yes

○ No

From Movement to Music, a Computational Creativity Approach

Please listen to the following videos.

Version 1

Version 2

Version 3

Which version did you preffer? *

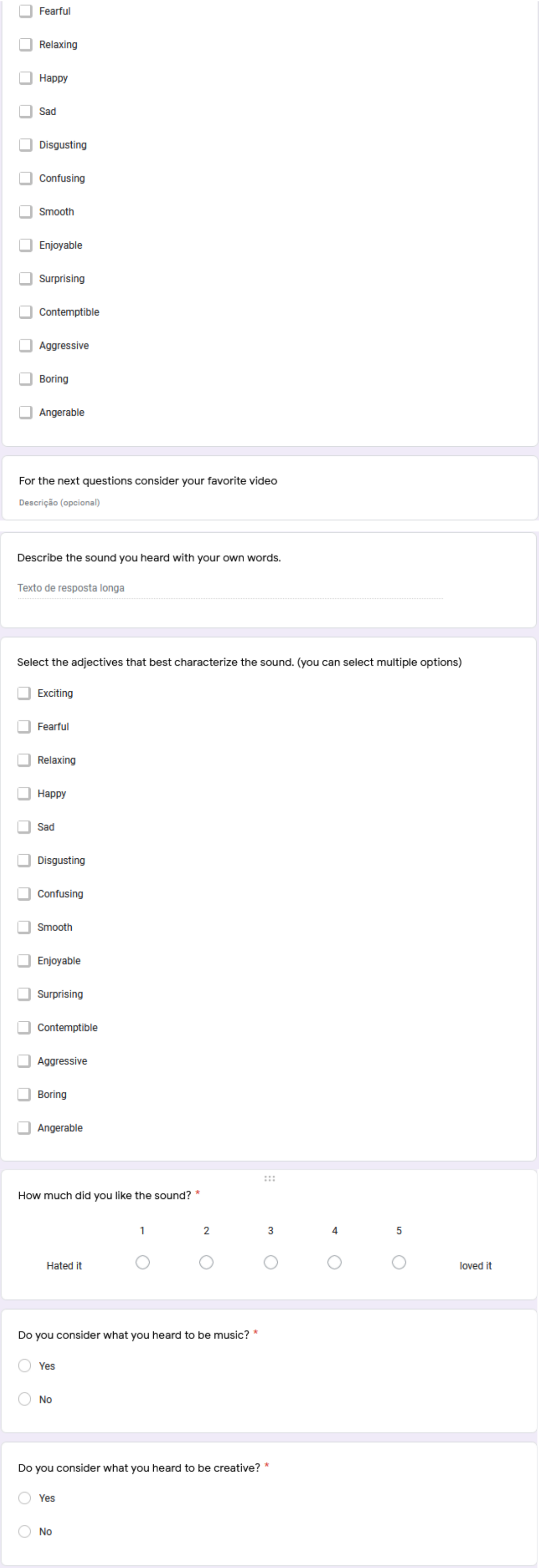
○ Version 1

○ Version 2

○ Version 3

Select the adjectives that best characterize the movement. (you can select multiple options)

□ Exciting



□ Fearful

□ Relaxing

□ Happy

□ Sad

□ Disgusting

□ Confusing

□ Smooth

□ Enjoyable

□ Surprising

□ Contemptible

□ Aggressive

□ Boring

□ Angerable

For the next questions consider your favorite video

Descrição (opcional)

Describe the sound you heard with your own words.

Texto de resposta longa

Select the adjectives that best characterize the sound. (you can select multiple options)

□ Exciting

□ Fearful

□ Relaxing

□ Happy

□ Sad

□ Disgusting

□ Confusing

□ Smooth

□ Enjoyable

□ Surprising

□ Contemptible

□ Aggressive

□ Boring

□ Angerable

How much did you like the sound? *

	1	2	3	4	5	
Hated it	○	○	○	○	○	loved it

Do you consider what you heard to be music? *

○ Yes

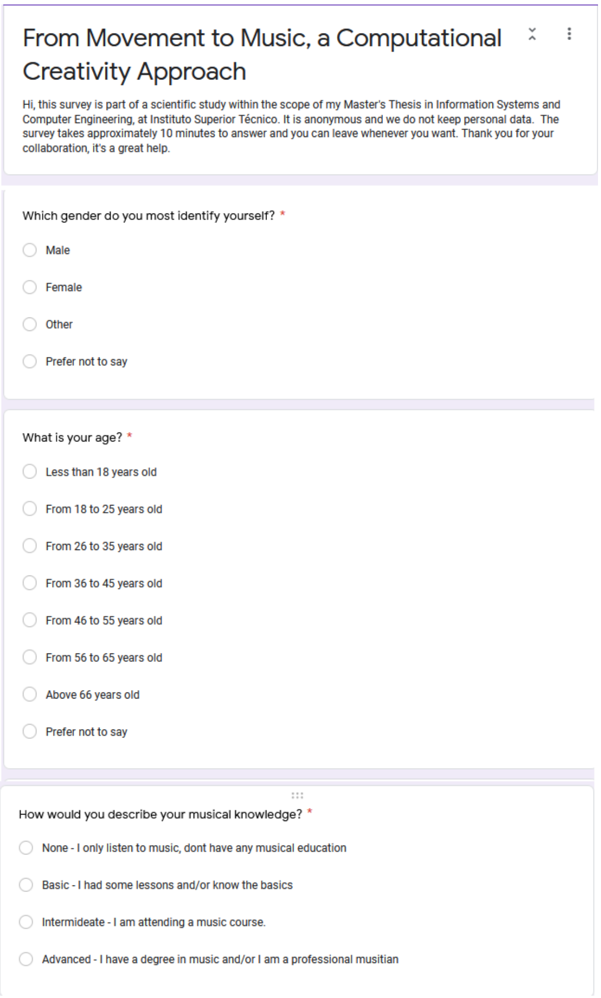
○ No

Do you consider what you heard to be creative? *

○ Yes

○ No

Figure B.1: Questioner-1



From Movement to Music, a Computational Creativity Approach

Hi, this survey is part of a scientific study within the scope of my Master's Thesis in Information Systems and Computer Engineering, at Instituto Superior Técnico. It is anonymous and we do not keep personal data. The survey takes approximately 10 minutes to answer and you can leave whenever you want. Thank you for your collaboration, it's a great help.

Which gender do you most identify yourself? *

○ Male

○ Female

○ Other

○ Prefer not to say

What is your age? *

○ Less than 18 years old

○ From 18 to 25 years old

○ From 26 to 35 years old

○ From 36 to 45 years old

○ From 46 to 55 years old

○ From 56 to 65 years old

○ Above 66 years old

○ Prefer not to say

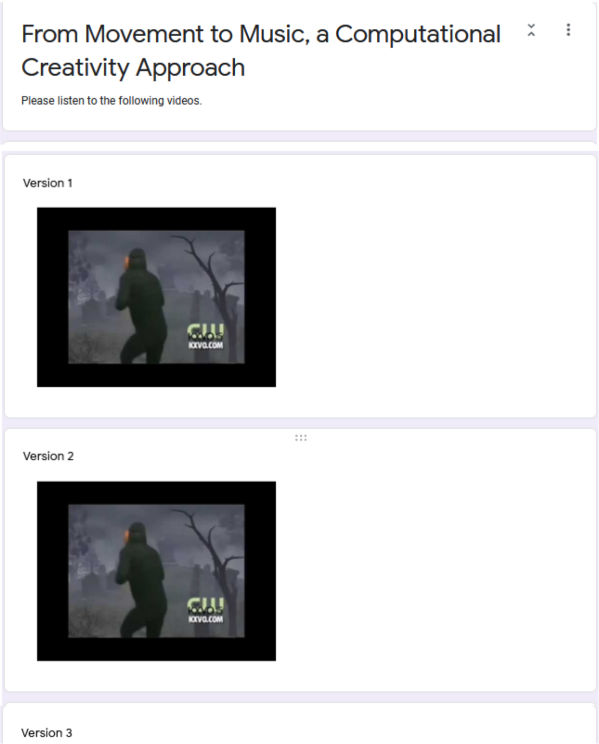
How would you describe your musical knowledge? *

○ None - I only listen to music, dont have any musical education

○ Basic - I had some lessons and/or know the basics

○ Intermideate - I am attending a music course.

○ Advanced - I have a degree in music and/or I am a professional musitian



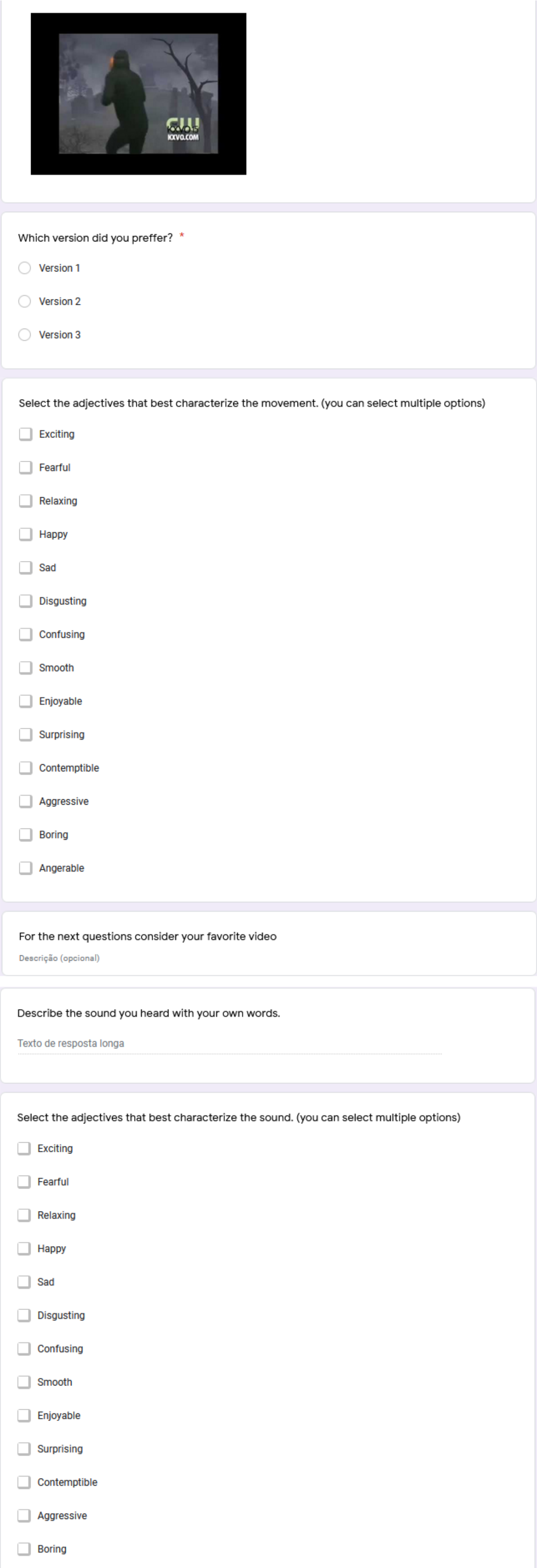
From Movement to Music, a Computational Creativity Approach

Please listen to the following videos.

Version 1

Version 2

Version 3



Which version did you preffer? *

○ Version 1

○ Version 2

○ Version 3

Select the adjectives that best characterize the movement. (you can select multiple options)

□ Exciting

□ Fearful

□ Relaxing

□ Happy

□ Sad

□ Disgusting

□ Confusing

□ Smooth

□ Enjoyable

□ Surprising

□ Contemptible

□ Aggressive

□ Boring

□ Angerable

For the next questions consider your favorite video

Descrição (opcional)

Describe the sound you heard with your own words.

Texto de resposta longa

Select the adjectives that best characterize the sound. (you can select multiple options)

□ Exciting

□ Fearful

□ Relaxing

□ Happy

□ Sad

□ Disgusting

□ Confusing

□ Smooth

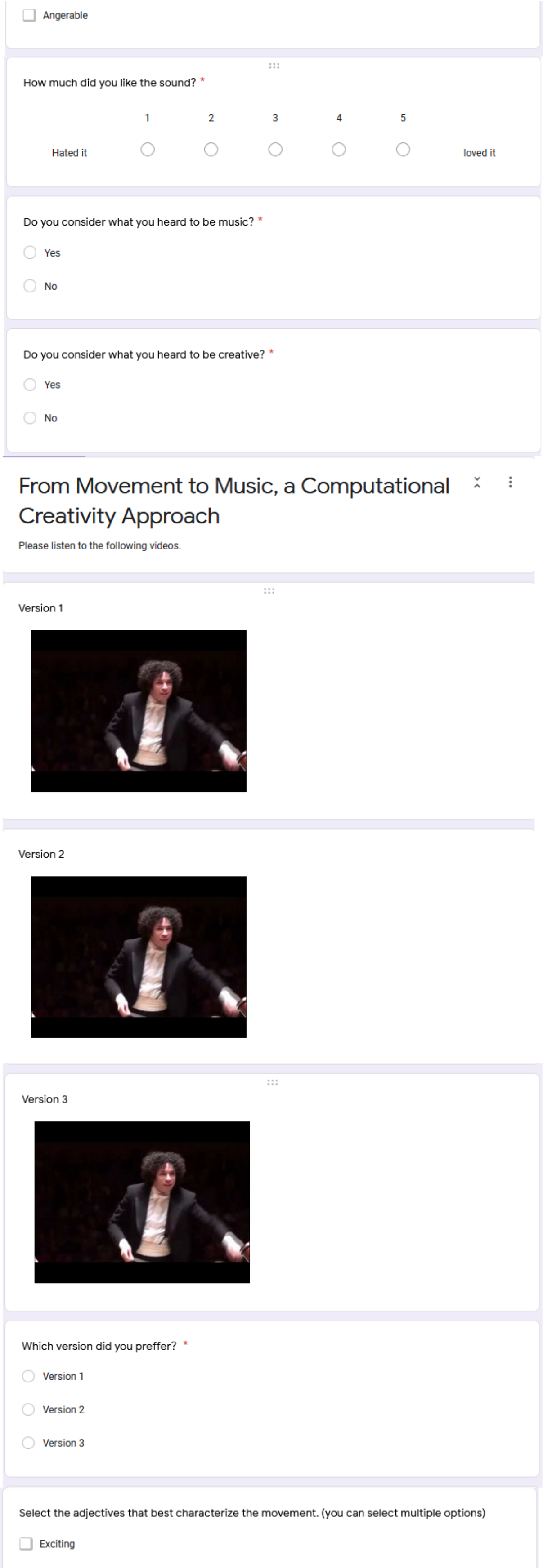
□ Enjoyable

□ Surprising

□ Contemptible

□ Aggressive

□ Boring



□ Angerable

How much did you like the sound? *

	1	2	3	4	5	
Hated it	○	○	○	○	○	loved it

Do you consider what you heard to be music? *

○ Yes

○ No

Do you consider what you heard to be creative? *

○ Yes

○ No

From Movement to Music, a Computational Creativity Approach

Please listen to the following videos.

Version 1

Version 2

Version 3

Which version did you preffer? *

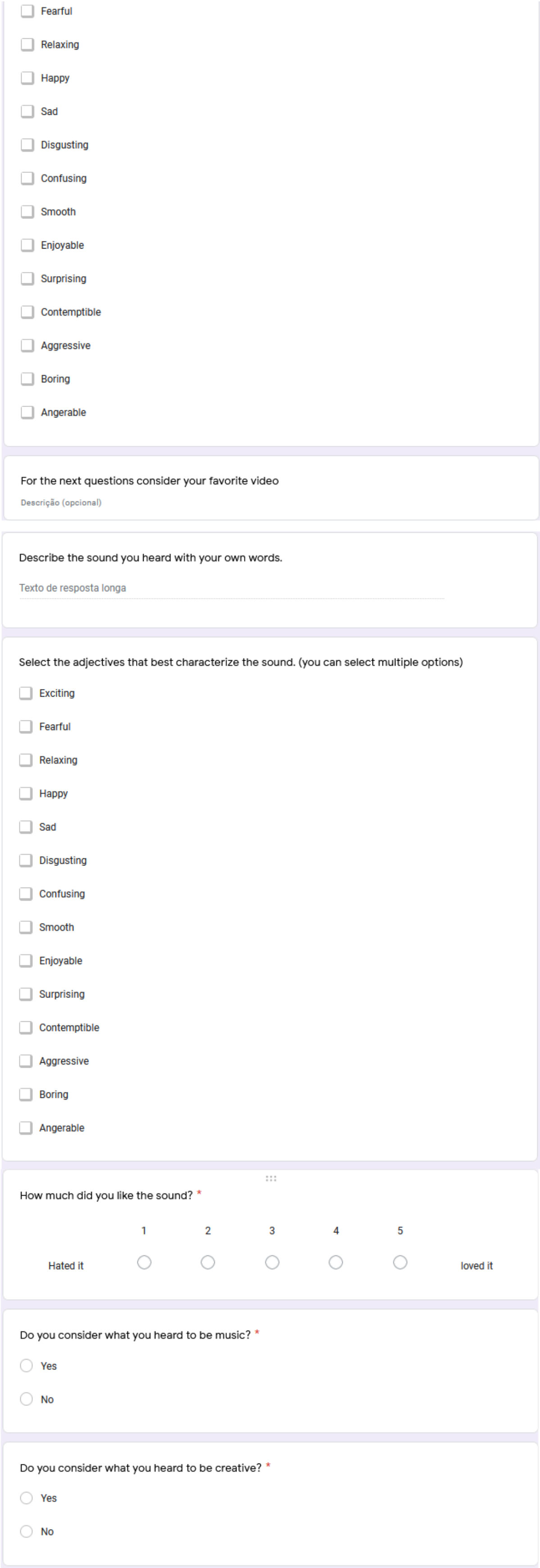
○ Version 1

○ Version 2

○ Version 3

Select the adjectives that best characterize the movement. (you can select multiple options)

□ Exciting



□ Fearful

□ Relaxing

□ Happy

□ Sad

□ Disgusting

□ Confusing

□ Smooth

□ Enjoyable

□ Surprising

□ Contemptible

□ Aggressive

□ Boring

□ Angerable

For the next questions consider your favorite video

Descrição (opcional)

Describe the sound you heard with your own words.

Texto de resposta longa

Select the adjectives that best characterize the sound. (you can select multiple options)

□ Exciting

□ Fearful

□ Relaxing

□ Happy

□ Sad

□ Disgusting

□ Confusing

□ Smooth

□ Enjoyable

□ Surprising

□ Contemptible

□ Aggressive

□ Boring

□ Angerable

How much did you like the sound? *

	1	2	3	4	5	
Hated it	○	○	○	○	○	loved it

Do you consider what you heard to be music? *

○ Yes

○ No

Do you consider what you heard to be creative? *

○ Yes

○ No

Figure B.2: Questioner-2