

AudioVisual Collage: A computational Framework For Audiovisual Composition and Computational Art

Diego Fagundes

PhD in Arts and Computational Technology

Department of Computing
Goldsmiths, University of London
Supervisor: Dr. Mick Grierson

London, United Kingdom
August 2017

Declaration

I declare that the work presented in this thesis is my own unless otherwise stated.

Diego Fagundes

Abstract

This thesis proposes a novel system for audiovisual composition called *AudioVisual Collage*. It consists of a screen-based interactive system, which contains a series of algorithms responsible for automating the tasks of image segmentation, inpainting, virtual camera motion, 3D lighting, visual transitions and visual motion generated from data extracted from audio files.

With the merging of theory and practice, this research project investigates new possibilities, new approaches, and alternative examples of what can be explored in the field of computational art so as to offer a computational framework embodying a set of conceptual principles for digital fine arts practice. The set of algorithms that constitute the AudioVisual Collage framework embody a collection of digital design techniques and artistic concepts created through an iterative, collaborative process with other artists, in which I acted as the sole computational practitioner. The two collaborative projects presented in this thesis, *Surround Sounds* and *Interactive Art Gallery*, are part of an on-going creation of case studies, which have allowed me to better understand a series of approaches and methods, leading to the design of a framework to successfully communicate and apply specific creative ideas.

AudioVisual Collage is a framework that allows other artists and designers to create their own audiovisual aesthetic experiences, generating rich media outcomes constructed by software in real-time. That is, the system incorporates approaches from traditional media, such as sound recording and cinema, in order to generate new computational media based on a set of clear design principles. It also proposes an art vocabulary for audiovisual composition in the digital domain, which aims to deliver something unique in terms of the inherent potential of digital work that is not possessed by other media.

AudioVisual Collage demonstrates an approach for manipulating the temporal structure and spatial organisation of visual elements so as to reflect their content, and also, a method for combining visual artworks and sound in the creation of algorithmic screen-based aesthetic experiences. Equally

important, it is in itself a statement of what could be possibly developed in terms of algorithms for audiovisual composition in the field of computational art.

Table of Contents

Chapter 1. Introduction	1
1.1. Contributions	4
1.2. Thesis Structure	6
1.3. Publications and Practical Work.....	9
1.3.1. Projects	9
1.3.2. Presentations	10
1.3.3. Awards	11
1.4. AudioVisual Collage Publication	11
Chapter 2. Literature Review	14
2.1. Design in the Digital Domain.....	14
2.2. Human and Non-Human Actors	16
2.3. Aesthetic Experience	18
2.4. The Perceptual Experience.....	21
2.5. Audio and Visual Information	23
2.5.1. Correspondence Between Audition and Vision.....	28
2.5.2. Auditory Perceptual Organisation	31
2.6. Closing Remarks.....	34
Chapter 3. Methodology	36
3.1. Exploratory Studies	40
3.1.1. LiveRoom - An Interactive Virtual Music Listening Environment.....	41
3.1.2. 12 Tones - Real-time Sound Visualisation for Live Performance.....	45
Chapter 4. Case Study 1 - Surround Sounds	48
4.1. Introduction	48
4.2. Project Development Process.....	51
4.3. Programming Motion.....	54
4.3.1. Extracting Data from Audio Files.....	55
4.4. Converting Audio Data into Motion	57
4.5. Creating Motion Randomness.....	59
4.6. Technical evaluation	61
4.7. Impact of Surround Sounds	62

Chapter 5. Case Study 2 - Interactive Art Gallery	64
5.1. Introduction	64
5.2. Project Development Process	67
5.3. Image Segmentation	69
5.4. Inpainting	71
5.5. 3D Composition.....	73
5.5.1. Virtual Camera	75
5.5.2. 3D Lights	76
5.6. The Role of Sound in the Interactive Art Gallery	77
5.7. Technical Evaluation	80
Chapter 6 - AudioVisual Collage	83
6.1. Introduction	83
6.2. Project Concept.....	85
6.3 Developing The First Prototype.....	87
6.4. Image segmentation algorithm.....	89
6.4.1. Thresholding Algorithm	90
6.4.2. Posterization Algorithm	92
6.5. Inpainting Algorithm	95
6.6. 3D Composition.....	99
Chapter 7 - User Study	103
7.1. Method	103
7.3. User Study Results	104
7.4. User Study 2	108
7.5. Discussion.....	109
Chapter 8. AudioVisual Collage Final Prototype	112
8.1. Using Brightness to Control 3D Lights	113
8.2. The Sound Engine of AVC	117
8.2.1. Extracting Data from Audio Files.....	118
8.2.2. Highest Amplitude Frequency Detect.....	121
8.2.3. Generating Motion from Audio Data.....	123
8.3. AudioVisual Collage Output	126
Chapter 9. Conclusion	135
9.1. Summary.....	135
9.2. Discussion.....	137

9.2.1. AVC Image Segmentation and Inpainting Algorithm	137
9.2.2. AVC Sound Engine Algorithm	139
9.3. Art as Research	140
9.4. The Art Studio as Laboratory	141
9.5. Future Developments	145
References	147
Appendix A [LiveRoom]	155
Appendix B [12 Tones]	159
Appendix C [Interactive Art Gallery]	164
Appendix D [User Study Questionnaire]	169

List of Figure

Figure 1.1. The artwork <i>Joobin</i>	11
Figure 1.2. Multiple views of the piece <i>Joobin</i>	12
Figure 1.3. The Invincibles Series presented at the Shirin Gallery in New York	13
Figure 3.1. Illustration of the practice-based research	37
Figure 3.2. LiveRoom implementation in PD-Extended	41
Figure 3.3. LiveRoom Music Listening Environment	42
Figure 3.4. Presenting LiveRoom at the 134th AES Convention in Rome, 2013	42
Figure 3.5. LiveRoom Graphical User interface	43
Figure 3.6. The crossfades control the audio level	44
Figure 3.7. 12 Tones Visual Display (Fagundes, 2014)	45
Figure 3.8. Presenting 12 Tones at the 137th AES Convention in LA, 2014	46
Figure 3.9. 12 Tones Graphical User Interface	46
Figure 4.1. <i>Surrounds Sounds</i> at the White Cube Gallery in London, 2015	49
Figure 4.2. A specific propriety of an audio file controls the position of a different layer.....	54
Figure 4.3. Converting audio into keyframes	55
Figure 4.4. Onomatopoeia scale being controlled by the audio amplitude	56
Figure 4.5. Using audio to control opacity	57
Figure 4.6. Onomatopoeia Y-axis scale controlled by the audio amplitude	58
Figure 4.7. The height value is being controlled by the data extracted from an audio file.....	59
Figure 4.8. The wiggle method accentuating the Y-axis movement	60
Figure 4.9. The wiggle expression	60
Figure 5.1. The overlapping of loops creates meaning in the IAG	65
Figure 5.2. Interactive Art Gallery Development Stages	66
Figure 5.3. Painting <i>Another Place</i> (Griffiths, 2014)	67
Figure 5.4. Painting <i>Vatican Slumber Party</i> (Lonsdale, 2014)	68
Figure 5.5. Elements extracted from the painting	69
Figure 5.6. Inpainted background image	71
Figure 5.7. Area Extracted in Image segmentation process	72
Figure 5.8. The Multiple layers of the piece <i>Another Place</i> placed within 3D Space	74
Figure 5.9. Objects' relative size changed by perspective	74
Figure 5.10. Camera Depth of Filed effect in the piece <i>Another Place</i>	75
Figure 5.11. Point light implementation	77
Figure 5.12. Sound Design of the piece <i>Another Place</i> (Griffiths, 2014)	78
Figure 5.13. Element highlighted in <i>Vatican Slumber Party</i> (Lonsdale, 2014)	79
Figure 6.1. Illustration of the practice-based research (Fagundes, 2016)	84
Figure 6.2. AudioVisual Collage visual processing (Fagundes, 2016)	88

Figure 6.3. Thresholding algorithm	90
Figure 6.4. The algorithm that calculates the image's brightness mean value	91
Figure 6.5. Adaptive threshold algorithm	91
Figure 6.6. Posterization algorithm applied	93
Figure 6.7. Posterization algorithm	93
Figure 6.8. Grayscale posterization algorithm	94
Figure 6.9. Posterization with Blur applied	94
Figure 6.10. Blur algorithm	95
Figure 6.11. AudioVisual Collage image segmentation and inpainting method	96
Figure 6.12. Turning an image section into a binary image	97
Figure 6.13. Original Image with brightness mean value of 138.4	98
Figure 6.14. Original Image with brightness mean value of 26.3	99
Figure 6.15. Images applied as textures to the 3D shapes	100
Figure 6.16. Images applied as texture to quadrilaterals generated within 3D space	100
Figure 6.17. Visual Layers placed within 3D space	101
Figure 7.1. Artworks produced by Participant 2, 8 and 1	109
Figure 8.1. 3D space with and without lighting effect	113
Figure 8.2. Image's brightness mean value controls the light intensity	114
Figure 8.3. 3D space with lighting effect applied	114
Figure 8.4. 3D lighting effect emphasizing the brighter background	116
Figure 8.5. Dialog box where participants introduce images and audio files	118
Figure 8.6. Introducing an audio file selected by the viewer into the processing chain.....	119
Figure 8.7. Audio recording of a Distant Rumble sound	120
Figure 8.8. Audio recording with highest amplitude frequency of 240Hz	121
Figure 8.9. Algorithm that identifies the audio frequency range with highest amplitude	122
Figure 8.10. Frequency response of the audio recording of a heavy rain	123
Figure 8.11. Information extracted from the highest amplitude frequency of an audio	123
Figure 8.12. Highest amplitude frequency in the Heavy Rain sound recording	124
Figure 8.13. Frequency response of a French horn single note	124
Figure 8.14. AudioVisual Collage visual processing	127
Figure 8.15. The frequency response of the sound recording of a car engine	128
Figure 8.16. Waveform of the sound recording of a warfare	129
Figure 8.17. AudioVisual Collage visual processing of Bombings.mp4	130
Figure 8.18. AudioVisual Collage visual processing of Pop.mp4	131
Figure 8.19. Frequency response of the audio recording of the wind	132
Figure 8.20. AudioVisual Collage visual processing of Sea_of_Fog.mp4	133

List of Tables

Table 4.1. The Table of Sounds displays the groups of sounds created by Marclay (2014c) and the proposed variations of motion generated from different audio recordings (Fagundes, 2014b)	53
Table 5.1. Table of Sounds recorded for the Interactive Art Gallery exploration of <i>Another place</i>	79
Table 8.1. List of outputs generated by Audiovisual Collage software	127

Chapter 1. Introduction

“The computer acts as a transitional object to mediate relationships that are ultimately between person and person”.

(Seymour Papert, 1980: 183)

This thesis details the creative and technical process underlying the development of a series of screen-based interactive artworks that have been created in collaboration with a number of artists, and exhibited nationally and internationally. These collaborative projects are part of an on-going creation of case studies, which have been designed to assist in the task of fully understanding a specific set of digital design approaches and methods for the purposes of creating digital computational art. The outcome of these projects support the creation of a framework that facilitates a specific set of creative approaches that constitute common elements, essential techniques and tools required, in terms of technology development, to support computational audiovisual composition in fine art contexts.

This research articulates a set of enquiries related to technical aspects of the collaborative projects developed throughout my PhD studies, so as to identify which of the development steps can be automated, with the creation of algorithms that can assist other artists to create their own screen-based artworks. This thesis includes a methodology chapter that describes my approaches across all practice-based work, outlining how I have engaged in the collaborative process throughout my PhD studies. In addition, it explains how these collaborations served as an instrument for the creation of a framework for screen-based audiovisual composition.

The questions this research tries to resolve are:

- In what ways can visual elements have their temporal structure and spatial organisation manipulated to reflect their content?
- In what ways can visual artworks and sound be combined for the creation of algorithmic screen-based aesthetic experiences?

The literature review focuses on key new media theorists in order to examine emerging directions in the language of moving images, including their temporal structure and spatial organisation, made possible through the application of computational approaches, (Murray, 2011), (Manovich, 2013), (Salter, 2010), (Hayles, 2012), (Boden and Edmonds, 2009), (Giannachi, 2012), (Goffey, 2008), (Cramer, 2005).

In the context of this thesis, the term temporal structure indicates the process in which the sound engine algorithm that is part of the practical outcome of this research manipulates elements extracted from still images to generate visual motion. That is, the platform combines and re-arranges audio and visual content input and outputs audiovisual material with a different spatial and temporal configuration, which consists of an on-going audiovisual aesthetic experience involving continuous change.

This thesis also includes literature on contemporary computational arts in the context of computer vision and audiovisual composition to observe the aesthetic concerns of current computer-arts practices (Wands, 2007), (Kwastek, 2015), (McLean, 2011). In order to discuss the possible methods for exploring audiovisual relationships, this research presents background and definitions regarding the auditory and the visual fields of study. Michel Chion's *Audio-Vision: Sound on Screen* (1994) provides a singular theoretical framework for studying the audiovisual relationship. Chion's contribution proposes terms such as *synchresis*, and explores others such as *acousmatic sound*, coined by the composer and *musique concrète* pioneer Pierre Schaeffer (1966), to expand the theory concerning sound and image.

The thesis presents two case studies that were created in collaboration with artists from different genres, which independently address specific

outcomes. Nevertheless, both projects share elements, techniques and approaches that can be automated with the creation of algorithms, and compiled as a design vocabulary, to assist other artists in the development of screen-based audiovisual artworks.

The term design vocabulary, in the context of this thesis, consists of a set of rules and constraints that outline a unique and consistent design style, which was created to guide the development of software for screen-based audiovisual composition. In other words, the design vocabulary presented as the practical outcome of this research consists of a set of rules for solving specific problems, in the field of computational art, as well as a creative activity, which can lead to new insights and new solutions to design issues. Besides, the design vocabulary presented in this research describes my choice of tools and approaches in the creation of software for computational art.

These projects form part of the iterative development approach that was central to the development of the framework, and associated software presented in this thesis. The practical outcome of this research, entitled: AudioVisual Collage, comprises an interactive platform containing algorithms that automate the key steps identified during development of the case studies. The software consists of a screen-based interactive system, which enables other artists and audiovisual designers to create their own audiovisual aesthetic experiences from a combination of digitised images and audio files of their choice. The software contains algorithms that automatically perform the tasks of image segmentation, inpainting, 3D composition, virtual camera motion, 3D lighting and conversion of audio data into motion. Each of these algorithmic approaches was specifically chosen and developed through case studies described in the thesis.

The practical portion of this thesis comprises the fundamental instrument and a means to effectively communicate the framework for screen-based audiovisual composition. The development of this research provides a tangible method for discussing the relationship between audio and visual information. Moreover, this thesis focuses on the mechanisms by which

technology can better enable and sustain the aesthetic dialogue between sounds and images in composition. The theoretical reflections presented in this thesis consist of an approach to the understand some of the ways in which human-computer Interaction and art can productively come together to help artists create more meaningful screen-based audiovisual work through generative means.

1.1. Contributions

The aim of this research is to effectively explore the creative process in which I took part as artist and programmer in collaboration with other artists. The study focuses on identifying my artistic decisions and the essential techniques that I utilised to create works of art in the digital domain. Hence, after isolating these techniques, they could be deeply analysed and automated as algorithms, which were combined to form a screen based interactive system, entitled: AudioVisual Collage. Through the process of evaluation, experimentation and testing, the practical portion of this research has been created with the precise purpose of allowing other art practitioners to create their own screen-based audiovisual artworks using computational technology. The main contributions of this research are:

- A method that combines automatic image segmentation and inpainting;
- A method for generating visual motion from audio data.

This thesis proposes a design vocabulary for art practice in the field of computational art. This approach has supported the creation of algorithms for audiovisual composition that includes elements from the sonic and visual disciplines. Nevertheless, the design vocabulary offered by this thesis remains relevant not only as elements of the software presented as the practical outcome of this research, but also as useful techniques to be implemented in other art projects.

This project demonstrates important techniques from auditory and visual disciplines. For instance, the debate surrounding audiovisual congruence (Chion, 1994; Grierson, 2005; Boltz, 2001; Abbado, 1988) presented in literature review chapter, which is vital for this investigation, has inspired and informed the creation of the algorithm to automatically perform the task of generating motion from audio data. Thus, this research has successfully produced a variety of associations that are illustrative of how phenomena present in the real world can be represented in the digital domain, and used to create art using computational technology. For instance, the algorithm presented as the practical outcome of this thesis, which is responsible for extracting data from audio files and use this information to generate visual motion, emulates the way in which objects vibrate while resonating to specific sonic frequencies in the real world. Another example of the associations that illustrate phenomena present in the real world is the effect generated by the algorithm responsible for 3D composition. To explain, the algorithm splits the image into three discrete layers, position them within 3D space, based on their brightness value, and applies motion blur to the layers. As a result, the focus point created on screen emulates the way our vision blurs elements of the field of view, which are not in our focus of vision.

The approach of producing digital simulations of phenomena present in the real world is significant to the practice and aesthetics of audiovisual computational art. This becomes even more important when we appreciate how deeply sonic and visual information are embedded features of our environment and essential for our perception of the environment (Boltz, 2001: 427). The practical portion of this research aims to present a design vocabulary for computational art practice by implementing simulations of physical aesthetic elements, while simultaneously developing algorithms to emulated elements from traditional media. In doing so, it explores the affordances of the digital domain to generate new proprieties that were non-accessible in those traditional media. This reveals the real *essence* of computational media, which is this continuous and stimulating construction of new techniques.

The practical outcome of this project rises from the combination of traditional media and the affordances of the digital domain. The algorithms presented as the practical portion of this research enables the creation of novel artworks by allowing the participants to input their own audio, and visual material, so as to create a piece of art that combines audio and visual data that have special meaning for them aesthetically. Thus, this research aims to demonstrate great creative potential with the implementation of a series of algorithms for computational art.

1.2. Thesis Structure

Chapter 1 (Introduction) introduces the thesis and present its style, aims, significance and impact in the field of computational art. It makes explicit the research project's goals and limitations, which have been recognised during the collaborative process that this thesis aims to outline.

Chapter 2 (Literature Review) comprises a review of the literature that builds the background and sets the context of this research. This chapter is divided into six sections as outlined as follows.

Section 1 (Design in an Evolving Medium) examines the modern history of the screen-based artistic practice in the digital domain and explores the intricate, and increasingly intimate, configurations of the human and the machine in the field of computational art.

Section 2 (Aesthetic Experience) offers a foundation for building a multidisciplinary approach to *aesthetics*, addressing the viability of perceptual aesthetics and discusses possible avenues of interest. It aims do generate a more holistic understanding, looking beyond the technological in order to explore the creative and experiential side of the digital domain.

Section 3 (The Perceptual Experience) acknowledges the correspondence between *aesthetic experiences* and the on-going audiovisual

exploration of our environment, combined with our own distinct preconceptions.

Section 4 (Audio and Visual Information) presents background and definitions regarding the auditory and the visual fields of study. This section explores the theory concerning sound and image so as to discuss possible methods for combining audiovisual material. Additionally, it investigates how audition and vision reinforce each other to create a coherent perceptual experience.

Section 5 (Closing Remarks) presents the final observations regarding the review of literature. This section summarises the central topics outlined in the previous sections of this chapter and indicates the way in which those concepts serve as important guides to the development of a framework for screen-based audiovisual composition in the field of computational art.

Chapter 3 (Methodology) presents a reflection on my methods across all practice-based work. This chapter also explains how the projects developed in collaboration with other artists served as a vehicle for the creation of a framework for screen-based audiovisual composition. Also, it describes a series of artworks that are directly related and contributed to the development of this thesis.

Chapter 4 (Case Study 1) presents the first collaborative project of the series, which consists of a multiscreen animation entitled *Surround Sounds*, presented by visual artist Christian Marclay (2015), in which I have worked as animator and audiovisual designer for a period of eight months. The project consists of a room-sized immersive video installation, in which onomatopoeic words, taken from the pages of vintage comic books, were animated and projected across four walls of a large room, from floor to ceiling. The central idea of the project was to translate sound into motion. Therefore, this chapter summarises the main techniques and approaches that I have developed in collaboration with Christian Marclay, to animate onomatopoeic images and to

make them behave, in terms of motion, in a manner that resembles or evokes their sounds; while the installation remains entirely silent. The project was presented during Christian Marclay's solo exhibition at the White Cube gallery in London, in January 2015.

Chapter 5 (Case Study 2) presents the second collaborative project of the series, entitled *Interactive Art Gallery* (Fagundes, 2014). It consists of a framework for combining visual artworks, music, sound design and non-linear narratives in the creation of screen-based aesthetic experiences. The development of the *Interactive Art Gallery* makes use of several techniques, including: extraction of geometric figures from an image's background; reconstruction of images with nearby textures; positioning of images and objects in 3D space; application of depth of field blur; implementation of 3D lighting and camera animation; music composition; and narrative design. This chapter describes its creative process and identifies essential techniques used. The framework was developed through a process of iterative design in collaboration with three visual artists, during which a series of four screen-based artworks were produced and delivered to the Debut Contemporary Art Gallery in London. The project was presented at the 136th Audio Engineering Society Convention in Berlin, in April 2014, and received the *Gold Award for Best Student Design Project*.

Chapter 6 (AudioVisual Collage) This chapter describes the project's development techniques and design principles, and outlines all the common elements and features identified during the collaborative process.

Chapter 7 (User Study) presents a user study implemented to generate an external and qualified view on the system's usage, aesthetics, interface and content. The feedback gained during this study was used to support the development of new iterations of the first prototype. That is, the results from the user study were fed back into the software's development process for the creation of the final prototype.

Chapter 8 (AudioVisual Collage Final Prototype) presents the primary practical outcome of this research project. This chapter explains how the AudioVisual Collage prototype was improved and how the sound engine was developed and incorporated into the final project.

Chapter 9 (Conclusion) presents the summary, contributions and conclusion of my practice-based research, reflecting on the combination of audiovisual composition and programming as a means for artistic expression. This chapter outlines how the projects developed during the collaborative process served as the basis for the creation of a framework for audiovisual composition in the field of interactive art.

1.3. Publications and Practical Work

1.3.1. Projects

2014

- Interactive Art Gallery (for Debut Contemporary Gallery) - Multimedia Web Application
- 12 Tones (with artist Marcos Fagundes) - Sound Visualisation for Live Performance
- The Real Me (with artist Sassan Bakhtiar) - Audiovisual Installation

2015

- Surround Sounds (with artist Christian Marclay) - Multi-screen animation
- Zoom Zoom (by Christian Marclay) – Photo Visualisation for Live Performance

2016

- The Invincible Series (by Sassan Bakhtiar) – created using AudioVisual Collage Software

1.3.2. Presentations

- Exhibition of “Interactive Art Gallery” at Dom Diogo de Souza Museum (Bage, Brazil, December 2013).
- Presentation of “Interactive Art Gallery” at the Arts and Computational Technology Seminar at Goldsmiths University (London, UK, February 2014).
- Presentation of “Interactive Art Gallery” at the 136th Audio Engineering Society Convention (Berlin, Germany, April 2014).
- PhD Portfolio presentation at Goldsmiths University (London, UK, September 2014).
- Exhibition of “The Real Me” audiovisual installation at Debut Contemporary Art Gallery (London, UK, 23rd to 25th of October, 2014).
- Presentation of “12 Tones” at the 137th Audio Engineering Society Convention (Los Angeles, USA, October 2014).
- Presentation of “12 Tones” at IRCAM Forum at Centre Pompidou (Paris, France, November 2014).
- Exhibition of “Surround Sounds” at White Cube Art Gallery (London, UK, January 2015).
- Exhibition of “The invincible Series” at the Behnam Bakhtiar Fondation (Saint-Jean-Cap-Ferrat, France, August 2016).

1.3.3. Awards

- Gold Award for Best Student Design Project for “LiveRoom” during the 134th Audio Engineering Society Convention (Rome, May 2013).
- Gold Award for Best Student Design Project for “Interactive Art Gallery” during the 136th Audio Engineering Society Convention (Berlin, April 2014).
- Silver Award for “12 Tones” during the 137th Audio Engineering Society Convention (Los Angeles, October 2014).

1.4. AudioVisual Collage Publication

The first series of artworks produced using the AudioVisual Collage software, was released on the 11th of August 2016, through the website Sediton Art (2016), which consists of an online platform for distribution of art in digital format. The series of videos, entitled *Invincible [Video]*, was presented by French artist Sassan Behnam Bakhtiar (Figure 1.1).

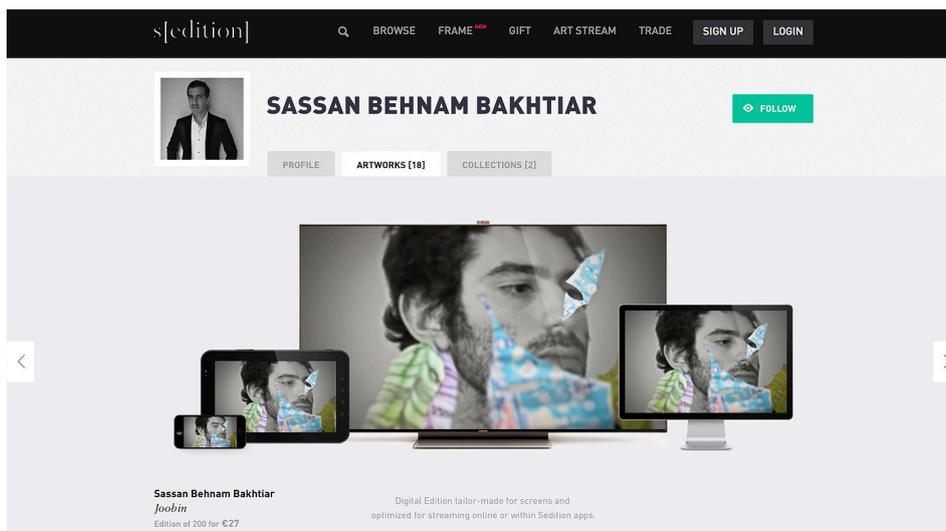


Figure 1.1. The piece *Joobin* (Bakhtiar, 2016) displayed at the Sediton website (seditonart.com, 2016).

The audiovisual artworks were created using the Audiovisual Collage software. According to Bekhtiar (2016), the series of artworks originated from the visual collages created by the artist combined with the soundtracks composed by Italian composer Pejman Tadayon. Both the images and audio files were input into AudioVisual Collage, which automatically generated the final artworks (Figure 1.2).

The Sedition website (2016) acknowledges the use of audiovisual collage in the creation of the video and outlines the aesthetic characteristics that were created and enhanced by the platform:

“Diverse animated motifs and patterns commonly seen in Iranian art and craft surround the image. The artist states: "There is a bit of everything in it, some is fictional, the rest is literally how I see that person in my head." The video work features the deep voice of Sassan himself, recounting the characteristics of Joobin as an empowering call to strength. The coloured tapestries overlaid over the black and white portrait, twist and flutter to a backdrop of traditional music by Pejman Tadayon with the animation created by the Audiovisual Collage software developed by Diego Fagundes at Goldsmiths University of London.”



Figure 1.2. The piece *Joobin* (2016) by artist Sassan Bekhtiar, which was created using AudioVisual Collage software (seditionart.com, 2016).

The static version of the *Invincible Series*, which comprises nine visual artworks heavily inspired by Persian mythology, are portraits of Iranian people, ornamented by coloured patterns of traditional Iranian tapestry. According to Bakhtiar (2016), the animation created by AudioVisual Collage software allowed the creation of an audiovisual experience that emphasised the most important elements present in the visual artworks. Besides, the audiovisual synchrony created a novel perspective of his own work by generating a sequence of rhythmic motifs and patterns, which follows the pieces of traditional Iranian music, which sets the tempo, rhythm and mood.

The series is currently being commercialised at the Sedition Art website. The series was also displayed in New York at the Shirin Gallery during the Future & Memory / Memory & Future Exhibition (Figure 1.3) between October and November 2016.



Figure 1.3. The *Invincibles Series* presented at the Shirin Gallery in New York (Bakhtiar, 2016).

Chapter 2. Literature Review

“Think of the computer not as a tool but as a medium”.

(Brenda Laurel, 1993: 21)

2.1. Design in the Digital Domain

One of the outcomes of the evolution of cyberculture is that our media history is constantly being re-configured from the new vantage points of the digital domain. As Lev Manovich states: “The narrative of media history is itself being amended by the present” (2002: 135).

The computer provides us with a new platform for the design of participatory experiences. We are progressively learning to act as performers, enacting emotionally genuine experiences that we recognise are not real (1997: 125). As Murray suggests “the more convincing the sensory representation of the digital environment, the more we feel that we are within the virtual world, and the wider variety of actions we will pursue to perform there (1997: 125). According to the philosopher Alva Noë, “the virtual presence is not merely illusory, it is real presence extended by new and different methods. The main representational property of the computer is the codified rendering of responsive behaviours” (2010: 84). That explains why we say that computers are interactive. They produce an environment, which is both participatory and procedural (2010: 84). From pioneer computer scientist Allan Kay’s perspective:

“The computer is a medium that can dynamically simulate the details of any other medium, including media that cannot exist physically. It is not a tool, though it can act like many tools. It is the first metamedium, and as such it has degrees of freedom for representation and expression never before encountered” (1984: 52).

In the same way that mathematics, printing and electricity revolutionised their own time in history, digital systems play a fundamental role in shaping the not only the material components of our workplaces and

homes, but also the structures of human culture. As researcher Lev Manovich suggests “the computer moved from being an enigmatic apparatus to become the engine of the future in only a matter of decades” (2008: 23). A substantial aspect that was directly responsible for the rapid popularisation of computers was the fast spread of software built specifically for art practices. For instance, word processors, hypermedia authoring, digital recording system, image editing and 3D modelling software allowed professionals used to utilised analogue methods, such as graphic designers, architects, engineers, filmmakers and musicians, to shift to new digital environment in order to solve the same tasks more quickly and efficiently (Manovich, 2008: 23).

Artists and researchers (Whitney, 1980; Cuba, 1983; Sorensen, 2014) have directly contributed to the advances in computer graphics, which is responsible for displaying visual information meaningfully and effectively to the user, revolutionising many types of media and significantly impacting animation, video games, cinema, advertisement and graphic design. Moreover, the advances in computer graphics have made a great impact on graphical user interfaces design, making the understanding and interpretation of data, as well as the interaction with the computer, much easier (Sorensen, 2014).

In his piece *3/78, Objects and Animations* (1978), animator Larry Cuba produced an analogy between the structure of a mathematical system and visual perception by means of computer language. Thus, proposing a novel approach for the organisation of aesthetic material, where sixteen visual objects, which vary between geometric and organic shapes, were programmed to perform rhythmic choreographies. More recently, computer animator Vibeke Sorensen has produced a piece called *Vishwaroop* (2014), which was programmed using Pure Data/GEM. The artwork consists of a real-time evolving system that constantly produces new elements and includes these forms as sources of data. Sorensen and Cuba’s artworks are great examples of new configurations that emerged between aesthetics and the digital medium.

One of the aims of this research is to rethink the intricate, and increasingly intimate, configurations of the human and the machine and to build a more holistic understanding, which follows the notion suggested by Lucy Suchman that “it is essential to look beyond the technological in order to explore the creative and experiential side of the digital medium” (1960: 24).

2.2. Human and Non-Human Actors

The design decisions that shape novel digital artefacts affect the way we think, act and communicate with one another (Murray, 2011: 8). Therefore, to design something new that transcends the expressive limitations of legacy media, and existing formats and genres, the designer should, think about the core human needs served by the new artefact. Grounded on that notion, it is significant to recognize that designers and artists should survey ways in which those needs could be met across multiple media platforms within the digital domain, and then, as Janet Murray suggests “attempt to reimagine them as they might be served by the affordances of the digital domain” (2011: 23).

The concept of *affordances* constitutes one of the central ideas on which this research is grounded. It was originally presented by the Psychologist James Gibson in his 1977 article *The Theory of Affordances and further explored in his book The Ecological Approach to Visual Perception* in 1979. Gibson defined affordances as: “all action possibilities latent in the environment”. In other words, the medium surfaces, substances, places, objects, and other beings have affordances for a given being, offering benefit or injury, life or death. The environment constrains what the being can do; therefore, the possibilities of the environment and the way of life of the being go together inseparably (Gibson, 1979: 143).

Cognitive scientist Donald Norman has appropriated the term affordances, in his book *The Design of Everyday Things*, in the Human-Computer Interaction Design context. The new appropriation of the term refers to those actions that are just readily perceivable by the actor. Thus, in the new context, the concept of affordances becomes dependent on the physical

capabilities of the actors as well as on their goals, past experiences and knowledge (Norman, 1988: 9). As Norman proposes “the height and the size of an object can readily indicate if a person could easily lift it or throw it, because it matches a person’s past experience with a similar object” (1988: 9). As a result, the new appropriation of term affordances is much more suitable for the solution of design problems. This research embraces the concept of affordances in order to guide the production of tools for audiovisual composition in the field of computational art.

It is important to understand that the computer was created not to transport static information but to embody complex behaviour (Murray, 1997: 72). Hence, as Murray points out “it is vital for the artist to orient creative design within the digital domain toward the underlying human needs, not the legacy media conventions” (1997: 72). As McLuhan enlightens: “We are never made aware of our environment until it becomes the content of a new environment” (1964a: 48).

It is essential to recognise the fundamental properties of the digital medium, that is, the features equivalent to the portability of the camera, the flexibility of the lens and the editability of film; that will define the power and form of a mature digital art (Murray, 1997: 68). Thus, since this research seeks to investigate how the elements that comprise the software that will be presented as the practical portion of this research relate to each other in aesthetic practices, it is relevant to introduce the concept of Actor-Network-Theory, first coined by Bruno Latour in 1980 and recently re-introduced in his book *Reassembling the Social* in 2005.

Actor-Network-Theory consists of an approach to social theory and research that according to Latour, “consists of an equivalent treatment of human and non-human actors within a network” (2005: 10). Therefore, it is an appropriate and useful tool for defining a process whereby software such as the one presented as the practical portion of this research can be developed. That is, in order to make sure the system advances successfully through the various stages of development, it is essential to verify both the cooperation of potential participants, as well as, the various components that constitute the

system. Further, since this research's outcome offers a series of algorithms built to facilitate other artists and designers to communicate and convey their artistic messages in the digital domain, it is essential to consider Latour's theory as part of this investigation.

When multiple elements are merged, we have a new form of agency, a new capacity for action, a new configuration that did not exist previously, which now enables new things to happen (Suchman, 2006). Bruno Latour exemplified the hybrid characteristic of a new actor-network configuration, with an illustration that he called *citizen-gun or gun-citizen*. In this context, he notes:

“You are different with a gun in hand; the gun is different with you holding it. You are another subject because you hold the gun; the gun is another object because it has entered into a relationship with you. The gun is no longer the gun-in-the-armoury or the gun-in-the-drawer or the gun-in-the-pocket, but the gun-in-your-hand, aimed at someone who is screaming. What is true of the subject, of the gunman, is as true of the object, of the gun that is held. A good citizen becomes a criminal, a bad guy becomes a worse guy; a silent gun becomes a fired gun, a new gun becomes a used gun, a sporting gun becomes a weapon (1990: 33)”.

Following Latour's theory, in the context of this research, all entities present in an audiovisual interactive system, including: the medium, the screen, the platform, the algorithms, the content, as well as the participants, should be recognised in equal terms. Therefore, in order to certify an effective and successful development process, this thesis presents a complete outline of its elaboration, as well as a user study that presents, the perspective of other artists while interacting with the system.

2.3. Aesthetic Experience

The investigation on aesthetic experience is relevant to this study as it observes art practice and participation through natural perceptual processes.

Since the practical outcome of this research consists of a tool designed to facilitate the creation of screen-based aesthetic experiences, it is important to acknowledge the correspondence between aesthetic experiences and the on-going audiovisual exploration of the environment, combined with the complex processes of perception. Therefore, the concepts presented in this section are expected to inform and guide the creative process through the biggest challenge it faces, which is to offer the participants a tool to allow them to create audiovisual aesthetic experiences.

The early twentieth century philosopher John Dewey, in his book *Art as Experience*, highlights that “artistic skill exemplifies skilful organisation of energy” (1934: 197). Therefore, great art finds and deploys ideal energy to create an experience envisioned to be experienced aesthetically. Every work of art, and every art discipline has a particular medium that it explores. Due to this, different disciplines reach the sense of totality in a given artwork, and the merging of ends and means, in qualitatively distinctive ways. For example, cubism is a good example of a powerful and skilful organisation of energy, as it comprises the means of seeing multiple aspects of an object from various points of view, simultaneously (Dewey, 1934: 197). Marshall McLuhan regards cubism as:

"the most radical attempt to stamp out ambiguity and to enforce one reading of the picture - that of a man-made construction, a colored canvas (1960: 26)".

The study of perceptual aesthetics seeks to trace the sense of pleasure achieved from art through natural perceptual processes. Perception being the on-going audiovisual exploration of our environment, combined with our own distinct preconceptions; and aesthetics being the mental and experiential pleasure achieved from art, which involves the whole brain, both emotional and cognitive pathways (Barry, 2006).

As John Dewey argues, in art and aesthetic experience, “means and ends coalesce, because means are ends in the aesthetics (1934: 36)”. In non-aesthetics, means are simply mechanical steps undertaken exclusively to

reach a desired end. That is, the non-aesthetics has a separation of means and ends; Dewey suggests the notion of *journeying* as an illustration. Non-aesthetic journeying is undertaken only to reach a destination. Therefore, any step that could shorten the journey is happily taken. On the other hand, aesthetic journeying is undertaken for the pleasure of traveling and contemplating what is around to be seen. Hence, extending non-aesthetic experience might lead to impatience and frustration, while prolonging aesthetic experience might intensify the sense of pleasure (Dewey, 1934: 37).

Following Dewey's (1934: 36) notion of aesthetic journey is very important for the development of this research, because it aims to generate an audiovisual journey proposed to be experienced aesthetically. That is, the system segments images in order to isolate elements of the image and apply motion to it, in order to reveal distinct aspects of the image. Therefore, this project aims to instigate the participants' pleasure of contemplating the aspects of the image that are being revealed.

As the American philosopher Alva Noë states, the meaning of art is related to how people experience it. Therefore, the biggest challenge faced by the artist is to offer the audience great experiences that are *aesthetically valuable*" (2006: 194). According to Noë, one of the reasons why art is so engaging and interesting to all, is that it recapitulates basic human elements such as: perception, understanding, skills, know-how, and the ability to bring the world into focus to make contact with it (Noë, 2006: 194).

Shimamura's research (2012: 21-22) contributes to the idea that numerous brain regions work together to process sensory signals and connect sensory information. Thus, it is essential to understand that many brain regions are involved in aesthetic experiences, and the main goal of perceptual aesthetics is to understand the dynamic interactions across these brain regions. Neuroscientist Ramachandran illustrates *perceptual grouping* principle with an example regarding the human ability to group colours; he points out that "vision might have evolved to permit human beings to navigate through intricate environments, identify dangerous situations, find nourishment, and to socially interact" (2011: 203). Ramachandran's research

suggests that in a scene where only several splotches of colours are visible, our brain is able to analyse the probability of these fragments to be exactly the same colour by coincidence and, if necessary, instantly join them together to identify and isolate the object. In sum, his contribution suggests that these continual cycles of iterative process help to disregard false solutions, in situations involving *noisy* visual images (Ramachandran, 2011: 203).

The concept of *aesthetic experience* will be used as a guiding principle for the development of this research. This is because it expresses the connection between art and perceptual experience and, in doing so, it underlines the relationship between the digital interactive system delivered by this research, the artists operating it, the artworks being created and the aesthetic experience it expects to generate.

2.4. The Perceptual Experience

The review on the topic of *perceptual experience* is essential for the development of this research as it proposes the notion that perception results from our interactions with the environment. Also, that the properties perceived through these interactions are not qualities of particular objects in themselves, but are qualities than should be attributed to human perceptual apparatus. This concept consists of a valuable knowledge to inform the development of the practical portion of this research as it aims to deliver artworks in the form of audiovisual experiences.

The central claim of what the philosopher Alva Noë (2006: 65) calls *Enactive Approach to Perception*, is that our ability to perceive is constituted by our possession of sensorimotor knowledge. Nevertheless, it is comprised not only by our mastery of patterns of sensorimotor dependence, but also by the fact that we *know* that our relation with the environment is mediated by such knowledge (Noë, 2006: 65). The essential idea is that perception is not merely having sensations; it consists of having those sensations and knowing what to do with them. Cognitive linguist George Lakoff's (2013a: 3) contribution points out that the word *environment* might suggest that it is

outside us, but in fact, it is also inside of us. We are connected to the physical world via our actions; we evolved to figure out how to act in the physical world. Likewise, Alva Noë emphasises that “the human understanding is not something autonomous or detachable. Rather, it is a skilful familiarity with and integration into the world. Just as people may count on their fingers and calculate using pen and paper, so they navigate within the world” (2010: 82).

The American neurobiologist Gerald Edelman (2007: 126) observes that the brain is embodied and, therefore, the brain and the body interact critically with each other. Moreover, both are embedded in the real world, which obviously has an enormous influence on their dynamics. For example, an object’s three-dimensional form can be seen and judged by the active observer, even if him or her are standing in a fixed position. That is possible because of our implicit understanding that *how things look with respect to shape* varies as a result of movement (Noë, 2006: 84).

In the process of perceiving, the brain facilitates a dynamic pattern of interaction between the body and the world. Experience is enacted by conscious beings with the help of the world (Noë, 2010: 47). We are not imprisoned in a dungeon of our own ideas and sensations. The phenomenon of perception, as life itself, is a dynamic and world-involving process. “We are already at home in the environment. We are out of our heads (Noë, 2010: Preface xiii)”.

Our description of the world, from a scientific point of view, is concerned with nature; our aptitude to create reflects our brain’s ability to give rise to a second nature. As Edelman points out: “Every act of perception is to some degree an act of creation, and every act of memory is to some degree an act of imagination” (2007: 100). In neurobiological terms, Edelman’s contribution (2007: 58-59) reveals that being selectional systems, brains operate at first glance, not by logic, but by pattern recognition. This is not a precise process, as is mathematics and logic. Instead, it trades off precision and specificity, if required, to expand its range. It is possible, for instance, that early human thought proceeded by metaphor, which, even with late

acquisition of particular means such as logic and mathematical thought, remains a key source of creativity and imagination in mature life.

Lakoff (2013a: 4) suggests that the concepts that rule our thoughts are not just related to our intellect. They also rule our daily functioning, down to the most ordinary aspects. According to him, our concepts function as a basic structure for what we perceive, for how we navigate around the world, and for how we socially interact. Therefore, our conceptual system plays a dominant role in defining our sense of reality. Following Lakoff's (2013a: 4) proposition that our conceptual system is generally metaphorical, then, it can be assumed that what we experience, what we do and the manner we think, is precisely a matter of metaphor. As Edelman states, "pattern recognition by the brain precedes logic, and early thought is creative in its pattern making through processes akin to metaphor" (2007: 65). That is, brain development depends on action in the world and, as a consequence, each brain is unique.

In a world that is perceived as a huge interconnected web, according to Edelman (2007: 46), it is essential to understand that we are not isolated viewers operating through *representations* in our mind. Instead, we are agents embedded in the world, learning through action in the environment. Besides, our brains are embodied and that embodiment is vital to support any understanding of how those brains work to gain knowledge. This framework of analysis is essential for this research as it helps in the understanding of the origin of perceptual categorisation, of concepts, and of thoughts based on interactions between the brain, the body, and the world.

2.5. Audio and Visual Information

The correspondences between audio and visual information are important guides to the development of the practical portion of this research. Therefore, it is essential to appreciate how perceptual system use information from one sensory modality to perform the perceptual organisation of the other one (Sekuler and Blake, 1985; Abbado, 1988; Bregman, 1999; Grierson, 2005).

American psychologist James Gibson, addresses an important question that is fundamental for the understanding of the relationship between audio and visual information:

“A fire is a terrestrial event with flames and fuel. It is a source of four kinds of stimulation, since it gives off sound, odor, heat and light. One can hear it, smell it, feel it, and see it, or get any combination of these detections, and thereby perceive a fire. For this event, the four kinds of stimulus information and the four perceptual systems are equivalent. If the perception of fire were a compound of separate sensations of sound, smell, warmth and color, they would have had to be associated in past experience in order to explain how any one of them could evoke memories of all the others. The problem of perception is not how sensations get associated; it is how the sound, the odor, the warmth, or the light that specifies fire gets discriminated from all the other sounds, odors, warmths, and lights that do not specify fire (Gibson, 1966: 54–55).

Gibson’s enquiry can be engaged in the following terms: How do the visual and the auditory systems blend their information to form audiovisual objects? Since, for instance, when a person speaks, it is not perceived as an auditory and a visual event, but as one. A well-known study by Sekuler and Blake (1985: 104) suggested that when a multisensory brain cell responds to a flash of light located in the upper right portion of the visual field, that same brain cell would respond to a sound only if it comes from the same location. Furthermore, when visual and auditory inputs occur simultaneously, a multisensory cell responds more intensely than when either input occurs alone.

Synchresis is a term coined by Michel Chion that, as he acknowledged, “is shaped from the combination of the words synchronism and synthesis” (1994: 63). It consists of the spontaneous and overwhelming blend generated

by auditory and visual stimulus that occurs instantaneously and, according to him, results independently of any apparent rational logic. The effect of synchresis is what makes dubbing and the post synchronisation process possible. Thanks to synchresis, for a particular face on screen, there are numerous acceptable voice sounds. In occasions where precise expectations of sounds are anticipated, synchresis is inevitable. For instance, the image of a person walking would accept almost any sound effect to sonically represent her footsteps. The synchresis effect allows filmmakers to make the most refined and astonishing audiovisual configurations (Chion, 1994: 63).

This idea contributes to this research as it aims to deliver tools for audiovisual composition in the field of computational art. That is, an important element of the software, which will be presented as the practical portion of this research, comprises a method for automatically synchronising audio and visual content.

A research study conducted by Aschersleben and Bertelson (2003: 157), asked participants to tap in time with a constant pulsation of flashes of light and disregard the sound that followed or preceded each flash. Although the instructions were in favour of the visual information, the participants' taps tended to gravitate towards the sounds. Once the visual and auditory pulse had their role reversed, and the participants were asked to keep in time with a pulsing sound, their taps gravitated towards the visual information much less than during the previous experiment. The experiment led the researchers to conclude that, the auditory system plays a greater role than the vision when processing temporal information.

As research by Chion suggests (1994: 11), visual perception and sound perception have their individual average pace grounded on their own very nature. The ears analyse, process, and synthesise faster than the eyes. According to him, an extremely fast visual movement does not form a clear figure, because its trajectory does not enter the memory as an accurate image. In the same length of time, a sound stimulus presents a more definite and clear form, which is recognisable, individuated, and distinguishable from others sounds. Chion goes on to explain this phenomenon: "sound is the

medium of language, and spoken words make the auditory system work really fast, in terms of processing time; in contrast, reading is notably slower” (1994: 11). That is, the eye perceives more slowly due to the fact that it has more information to process at once; it must explore in space as well as follow along in time. So, according to Chion’s notion, in a first contact with an audiovisual message, the ear is more temporally adept and the eye more spatially adept.

Adriano Abbado (1988: 5) proposed a framework to analyse the placement of audiovisual sources in space. He suggested that, generally, music should be treated as a form of art that explores not only the *time* dimension, but also the *space* dimension, in the same way as visual arts must be understood as involving *space* as well as *time*. According to him, this view has not been explored in any substantial way, and it has been used mainly as an effect, rather than as a definite musical parameter. Abbado’s framework is an important guide to the development of this research. To explain, the notion proposed by Abbado, in which the time and space dimension should be equally recognised and addressed in the musical context, will inform the development of a sound engine algorithm that will be responsible for extracting information from audio files to generate visual motion.

However, it is important to notice that sound localisation, cannot be as accurate as it is in the visual domain, and is essentially a function of the spectral content of the sonic material. This fact influences the association of shapes and sizes to sounds, in the sense that the size of a particular shape is a function of the spatial extension of the corresponding sound, which is in turn, a function of the spectral content of the sound itself. This audiovisual relationship is illustrated on an animation created by the Japanese artist Yoichiro Kawaguchi entitled *Ecology II: Float* (Kawaguchi, 1987). In this artwork, the author created a scale of sizes that, in turn, corresponded to a scale of sounds, from the lowest to the highest in frequency. The sound that contained the highest perceivable frequency was the smallest (precise localisation), while the sound with lowest frequency filled the screen, simulating its vague spatial localisation.

According to Chion (1994: 69), films present a peculiar sensory experience with regards to the localisation of sounds. The point from which a sound stimulus is produced, typically a loudspeaker, is usually not the same point on the screen where the sound is supposed to originate from. Nevertheless, the audience perceives the sound as coming from the visual source on the screen. It consists of a phenomenon that Chion describes as *spatial magnetization* of sound by image. For instance, when a character walks across the screen, the sound of her footsteps seems to follow her visual clue, even though they are being issued from a stationary loudspeaker located within the film theatre space. Likewise, if the loudspeaker is located relatively far from the screen, as in an outdoor setting, or if the soundtrack resonates in our head by means of headphones, these sounds will still be perceived as coming from the screen, despite the evidence of our own senses. Yet, as Chion explains, it is important to notice that a sound stimulus coming from another point than the screen is *magnetizable* only if the sound itself maintains a basic spatial stability. If the sound moves back and forth between loudspeakers, the image will have a harder time absorbing it, and the sound takes on a centrifugal force of its own that resists visual attraction. Since the practical outcome that this research expects to deliver comprises a platform for screen-based audiovisual composition, the concept of spatial magnetization of sound, proposed by Chion, is an important guide to the development of this research.

The work developed by Normal McLaren is another important guide to the development of this research. McLaren has developed several groundbreaking methods for combining and synchronizing sound and animation. McLaren began experimenting with animated sound in the 1930s, drawing and scratching on optic film. As part of his experiment, McLaren built a library of hundreds of hand drawn sound waves' patterns, which were able to generate all semi-tones in a range of five octaves. McLaren registered the images that produce each specific musical note onto optical film. Thus, the audience was able to visualise the soundtrack on screen. During McLaren's experiments with animated sound, he developed different types of sounds.

For instance, McLaren produced hard-edged squared stripes that generated square sound waves. Also, by slightly rounding the edges of the stripes, McLaren was able to change the audio response and, thus, the tone produced by the waves. McLaren's work illustrates approaches to perform audiovisual synchronisation. Thus, his work consists of an important guide to the development of this research.

As research by Noble (2012: 341) suggests, engineers, architects, interior designers, and most of all, musicians deeply comprehend that the way we perceptually experience the environment around us is shaped and guided by sound. Hence, sonic information can also be seen as an essential constituent of a complete and rich perceptual experience in an interactive system within the digital domain. Thus, the investigation on the relationship between audio and visual information are important guides to the development of this research.

2.5.1. Correspondence Between Audition and Vision

Researcher Albert Bregman (1999: 201) proposes, in his book *Auditory Scene Analysis: the Perceptual Organization of Sound*, a series of Gestalt principles for auditory scene analysis that present similarities between vision and audition. According to his contribution, principles of auditory perceptual organisation, such as similarity (the grouping of elements), common fate (the segregation of a group of elements) and good continuation (the continuity of hidden or overlapping elements) appear to perform similar roles in both sensory modalities.

The audio engine algorithm that this research expects to deliver is anticipated to generate visual motion by extracting data from audio files. The system is expected to produce audiovisual aesthetic experiences that will explore the connection between the images and the sound files input by the participants. Thus, the investigation on the concepts of perceptual organisation of sound and the principles that govern the correspondence

between audition and vision is essential to inform the development of this research.

A study conducted by Robart and Rosenblum (2005: 164), points out the directivity and intensity of an incident light enables the visual system to process the properties of a visual object. The information available to the auditory system is simultaneously processed, and utilised to characterise the dimension of the physical space in which a sound is produced. Their study suggests that the auditory system's main purpose is to recognise permanent properties of objects, such as the nature of sounds produced by their own activity, if they are animated, or the audible response to physical energy if they are static. The above considerations indicate that regarding the concepts of source and surface, audition and vision reinforce each other to create a coherent perceptual experience. One of the

A contribution by Henrisken (2002: 34) suggests that the recognition of a sound source is, to a certain degree, influenced by our knowledge and expectations. When one listens to a familiar sound, this process occurs unconsciously and at a high speed. But, when listening to new sounds, a person becomes aware of the recognition process. The decision whether to accept that she cannot detect the sound source, or to carry on an investigation, suddenly becomes a conscious one. Henrisken also underlines that "the other senses play a significant role in the auditory recognition process" (2002: 114). For instance, visualising the sound source as it produces the sound, or producing a sound by direct bodily interaction, such as playing a musical instrument, adds further knowledge about the sound source.

In contemporary society, the human sensory system is continuously flooded with images and sounds combined to draw attention, influence behaviour and manipulate state of mind. According to Lipscomb (2005: 65), one of the most effective uses of audiovisual material, specifically intended to trigger perceptual response, is found in screen-based aesthetic experiences, such as motion pictures and animation. As researcher Grierson (2005: 18) elucidates, synchronised sonic and visual material can radically change the way in which the audience perceives a given visual object on screen. Thus, it

generates a third *audiovisual* form. Heavily synchronised audiovisual material is a very effective approach used for the creation of aesthetic identities. According to him, this technique is commonly utilised by advertisers and media companies through the combination of graphic material with sound effects and music.

Pierre Schaeffer (1966: 91) coined the term *acusmatic sound*, drawing upon the *acusmatic* mode of listening proposed by the Greek philosopher Pythagoras. This term refers to the teaching practice that consists of addressing appendices through a curtain, not allowing them to see the lecturer. In other words, *acusmatic sound* is an auditory stimulus that one hears without seeing its originating cause. Telephone, radio and phonograph, all which transmit sounds without showing their emitter, can be understood as *acusmatic media*.

Michel Chion (1994: 71) investigates *acusmatic sound* from a cinematic perspective, to emphasise a deeper level of significance of this term, and to stress the fact that it is more a reasoning than a situation. As he suggests, in cinema, the *acusmatic* situation can emerge in two different ways: the sound source can be first seen and later *acusmatised*, or conversely, the sound can be initially *acusmatised* and have its source revealed subsequently.

The first scenario allows the association of a sound stimulus with a specific image, what Chion (1994: 72) calls *visualised sound*, and Schaeffer (1966) refers to as *direct sound*. In this case, according to research by Chion, it becomes an embodied sound, identified with an image and, therefore, classified. In the second instance, the sound source remains veiled for some time, to intensify tension, and is only later revealed. This has the effect of *de-acusmatising* the originally hidden sound source; a cinematic feature that is frequently employed in suspense and mystery films (1994, 73).

Another significant element in the analysis of audiovisual relationships is the notion of *audiovisual congruence*, and how it affects our memory. Research conducted by Marilyn Boltz (2001: 427) suggests that the nature of a sound can directly affect the memorability of a visual event. The memorability of the audiovisual combination as a single entity will depend,

ultimately, on the level of congruence between the audio and visual material at the moment they are synchronised.

The similarity between the structures of an image and a sound stimulus relies entirely on the structural aspects of their components. As Grierson (2005: 19) points out, a sound that has a high ratio of unconnected overtones to harmonic ones can be described as *noisy*. Similarly, an image that exhibits less self-similarity, less order and more incoherent complexity, could also be described as noisy. This concept offers a tangible method for discussing audiovisual congruence, as we are able to contemplate sounds and images in a similar fashion regarding their texture (material, shape, size), and timbre (tone quality and frequency response).

2.5.2. Auditory Perceptual Organisation

According to Albert Bregman (1990: 203), auditory perceptual organisation consists of the process of partitioning and grouping acoustic elements of a sonic impulse. The investigation on auditory perceptual organisation is central for the development of the practical portion of this research. This investigation outlines principles that will be used to inform the development of an algorithm responsible for the segregation of certain elements of sound files from others.

As suggested by researcher O'Callaghan (2009: 2), sounds are the means that carry the audible qualities that endure changes and persist over time. Accordingly, sounds can be comprehended as singular entities, instead of merely repeatable properties. According to O'Callaghan, this unrepeatability feature found in sound, characterise it as an *event* like object. Sounds can be viewed as a certain type of event that can be described as a sound-event, rather than simply a feature that a visual object holds. Since sound-events happen in a certain place, in a certain time, in a certain way, it is reasonable to say that they are *objects* of auditory experience.

The ability to form auditory objects is an essential one in the physical environment, because sounds that arrive at our ears might have arisen from

distinct auditory events. Furthermore, humans rely on auditory grouping mechanisms to communicate, particularly in the presence of outer noise and competing sonic events. These multiple perceived organisations become acknowledged units, and once they gain our awareness, they are perceived as objects. That is to say, audition and vision begin to pull together information that is not likely to be accidental; and whether the process of perceiving objects is affected by attention, or occurs instinctively, it is now isolated from all other entities. In his contribution, *Perceptual Correspondences of Abstract Animation and Synthetic Sound* (1988), researcher Adriano Abbado investigates the concept of aural objects. He explains that in the same way that shape and surface define a visual object's identity, the sound object is defined by its sonic spectrum. To put simply, shape and surface attributes in images, and timbre in sound are the greatest perceptual determinants.

As proposed by British cognitive scientist Colin Cherry (1953: 975), the same principle can be applied to the auditory system. It consists of a phenomenon he described as *Cocktail Party Effect*, in which our brain analyses a scene, by perceptually organising sensory data, to form sound objects (or auditory streams) allowing people to *tune into* a single voice and *tune out* all others. To explain, a person analyses an auditory scene grouping relating components from one sound source, and segregating components (and frequencies) that are not part of the attended sound object. In his seminal book, Albert Bregman describes auditory stream segregation as:

“The general process of auditory scene analysis in which links are formed between parts of the sensory data. These links will affect what is included and excluded from our perceptual descriptions of distinct auditory events” (1990: 47).

The researcher Albert Bregman's suggests in his book *Auditory Scene Analysis* that the word *sound* is usually used to refer to both the physical sound stimulus and to our perceptual experience of it. However, it is essential differentiate them by using the word *stream* to describe the perceptual

experience of sound, and the word *sound* or *sound-event* to describe its physical cause (1999: 10). *Auditory scene analysis* can be defined as the process of distinguishing *streams* during the occurrence of intrusive noise. *Streams*, in turn, consist of the place for attachment of the specific sonic qualities, changes and co-existence with other simultaneous streams (O'Callaghan, 2009: 2). Based on that notion, it is evident that we group different sound-events, isolating them into separate auditory streams (Chrysakis, 2010). As Bregman suggests (1999: 456), during music listening, distinct organisational dimensions take place, from the analysis of melody to the analysis of timbre, pitch, sonic textures, depth, directionality and polyphonic layering.

As Bregman suggests (1999: 457), the principles that control auditory scene analysis in natural environments generate perceptual representations of the separate sound-events perceived in that environment, such as: the wind, the rain, voices or a vehicle approaching. In the musical context, there are individual events too. However, these representations do not always characterise basic units of experience. That is, through a piece of music, the artist often attempts to fool the auditory system by generating fictional streams. For example, different music instruments might be used to play a chord at the same time so as to be perceived as a whole.

A fundamental process in the music listening process is the approach employed by listeners in order to select certain elements among the abundant amount of information that emerges from a musical piece sonic surface. As listeners perceives, their attention oscillates among the several pathways of this sonic surface. The information that is retained and stored for later assessment is dependent on the type of information attended during the listening process (Chrysakis, 2010: 56). The concepts presented in this section function as guides to the development of this research as they inform ways in which audio and visuals can be explored, combined and re-arranged to generate new audiovisual content.

2.6. Closing Remarks

The literature review chapter has explored the concepts of aesthetics, perception and the audiovisual field of study so as to contemplate the creative and experiential side of the digital domain as well as the aesthetic concerns of current computer-arts practices. Also, this chapter indicates the way in which those concepts serve as important guides to the development of a design vocabulary for screen-based audiovisual composition in the field of computational art.

My analysis, in sum, is that the artistic practices in digital domain must be able to offer something unique in terms of some inherent potential not possessed by other media. Computer software can fulfil this condition by creating in a systematic and convincing way, powerful aesthetic experiences, underlying properties of the natural world as well as the subtlety of human's relationship with it.

Grounded on this notion, this research project seeks to resolve two main questions. First, in what ways can visual elements have their temporal structure and spatial organisation manipulated to reflect their content? Second, in what ways can visual artworks and sound be combined for the creation of algorithmic screen-based aesthetic experiences?

Therefore, the review of the literature is important for the development of this research project as it describes new directions in the language of moving images made possible through the application of computational approaches, (Murray, 2011), (Manovich, 2013), (Salter, 2010), (Hayles, 2012), (Boden and Edmonds, 2009), as well as, investigated methods for exploring audiovisual relationships, which included the definitions regarding the auditory and the visual fields of study (Chion, 1994; Bregman, 1999; Schaeffer, 1966), to expand the theory concerning sound and image. Thus, it serves as both a guide to the development a framework for screen based audiovisual composition as well as a guide to what is being developed on the contemporary computational arts field in the context of computer vision and audiovisual composition to observe the aesthetic

concerns of current computer-arts practices (Wands, 2007), (Kwastek, 2015), (McLean, 2011).

Chapter 3. Methodology

“The details are not the details. They make the design”.

(Charles Eames)

This project consists of practice-based research. Accordingly, it comprises an original investigation undertaken to gain knowledge by means of practical process, as well as, by the outcome of the practice. The practical portion of this thesis, which consists of a screen based audiovisual composition tool, entitled: AudioVisual Collage, along with the collaborative process with artists and the outcomes of these collaborations consist of instruments to communicate a framework for screen-based audiovisual composition.

This relationship between software and art leads to a broad conversation related to current practices, which include: interactive installations, generative art, net art, video games, browser art, creative programming, and so on. This thesis is less focussed on defining all current phenomena and practices within software art. Nevertheless, the methodology used in this research comprises a process whereby through collaborations and free form artistic practice a series of techniques and approaches that could be automated were identified. The software presented as the outcome of this research captures my artistic approaches. The main achievement this thesis expects to accomplish is to make these artistic approaches available to other artists and designers. As a result, this thesis focuses on the aesthetic processes, which are triggered by programme code.

This project is driven by a combination of theory and practice, to provide a unique design vocabulary for audiovisual composition in the field of computational art. Firstly, the review of literature provides a general view of the problem area outlined in the introductory chapter. The literature review chapter describes new directions, in the language of moving images, made possible with the application of computational approaches in the contemporary computational arts field. Also, it explores the aesthetic concerns

of current computer-arts practices, in terms of computer vision and audiovisual composition methods.

Secondly, this thesis documents the creative process and presents an analysis of two collaborative projects, in chapters 4 and 5, to specifically identify the steps of the creative process that could be automated with the creation of algorithms to allow other artists and designers to create their own artworks. Thirdly, based on the results from the collaborative process analysis, the first prototype, which comprises *image segmentation* and *inpainting* algorithms, was developed and evaluated. The evaluation of the prototype was done through the creation of a user-study, which applied a usability evaluation methodology so as to collect feedback and insights from other artists and designers in order to better inform the final design. The main user-study methodology was complemented by a second study, which introduced the participants with an additional task, to understand their performance and usability needs while operating the system.

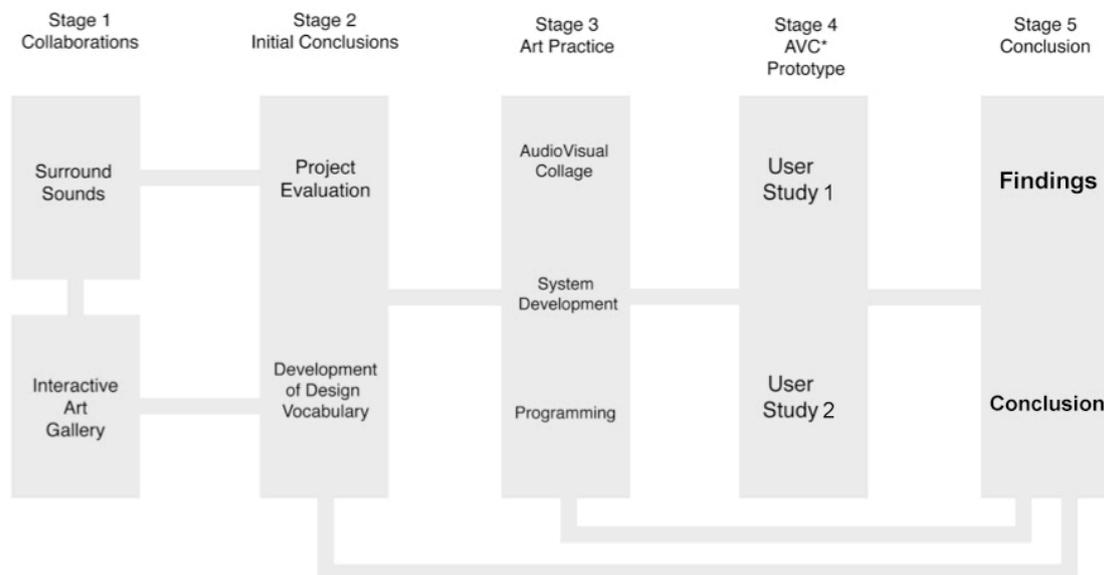


Figure 3.1. Illustration of the practice-based research (Fagundes, 2016).

The study was conducted to obtain the observations and opinions of other visual artists and designers after interacting with the system. Likewise, the study was created to discover if other artists would need or want such a tool. Since the aim of this thesis is to deliver a design vocabulary in the form

of a series of algorithms for art practitioners working in the field of computational art, it was vital to obtain their evaluation with regards to the system's overall structure and features (Figure 3.1).

The methodology of this thesis follows the notion advocated by Höök et al (2003) in the publication entitled '*Sense and Sensibility: Evaluation and Interactive Art*', which outlines the importance of user experiments for improving the design of computational interactive systems. As the study suggests, users' evaluation can help fine-tuning the design of a system, as well as, to ensure the interaction works as intended. The study also points out that user experiments can expand the understanding and logic, which in turn, can be applied to build a more stable system and meet the initial creative vision. Finally, the outcome of this evaluation was fed back in the creation of the final version of the system.

This research project was developed through an iterative process. Thus, different ideas were constantly being tested and evaluated so as to reveal alternatives to traditional approaches and conventional paradigms, and to support the creation of audiovisual aesthetic experiences. According to research by Edmonds (2005), iterative process of development, where new ideas and insights are rapidly fed back into the development process, supports and enables the development of new forms and knowledge, creating a circumstance that enhance development possibilities. Moreover, iterative design is a form of practice-based research that explores existing technology in new ways, which can lead to the transformation of existing and traditional practices. As research by Dix et al. (2004) suggests, when the artist has completed a sufficient number of development iterations, the specification of a prototype version can be formulated and implemented for evaluation. Later, the results from the evaluation are then fed back into the development process for the creation of a final prototype (Dix, Finlay, & Abowd, 2004).

The two collaborative projects previously mentioned are part of a continuous creative path that culminates in the screen based interactive system: *Audiovisual Collage*. Although the two collaborative projects address

different specific outcomes, they share various techniques and approaches. Hence, it is possible to detect the similarity between them.

This research project investigates new opportunities of what could be explored and proposes a novel approach to the field of computational art. Likewise, the practical portion of this research is envisaged to be in itself a statement of what could be possibly developed in terms of algorithms for audiovisual composition and, equally important, as my own statement as an art practitioner working in the field of computational art.

The creation of aesthetic experiences is a fundamental concern of this investigation. Therefore, in order to tackle the issues of aesthetics in practical terms, this project proposes algorithms that mimic processes present in the real world (eg. lighting, motion, perspective) and deal in a holistic way with complex processes, including experience, perception, participation, representation and understanding.

The AudioVisual Collage software consists of a system centered in experience. That is, the significance of the system lies in the active and dynamic process of aesthetic experience, which includes the audiovisual artists and designers operating the system as well as the audience experiencing the outcome. As a result, the system could not be evaluated using quantitative methods, as the measure of its effectiveness lies on the perspective of proficient participants, for whom the system was envisaged when developed.

The user study focuses on assessing whether the research questions are being addressed by the early prototype. Also, the evaluation aims to gain feedback, from the perspective of professional audiovisual practitioners, with regards to outcomes that had not been expected. The final outcome of the user study is a valuable indication to whether the initial ideas and concepts have been correctly implemented in the first prototype.

The *AudioVisual Collage* software is arguably a novel approach to audiovisual composition. Therefore, the focus of the study is to present a prototype to professional art practitioners so as to gather information to achieve the right design. The survey aims to recognise positive and negative

aspects of the project, as well as opportunities and areas of interest that could be further developed and refined.

This research project presents contribution to knowledge through the creation of a design vocabulary for screen-based audiovisual composition in the field of computational art. This design vocabulary is merged and implemented, in the construction of the practical portion of this thesis, to assist other artists and designers to create their own aesthetic experience. Additionally, the software which will be presented as the practical outcome of this research functions as my own artistic statement, as it comprises a series of algorithms that mimic and automate the essential artistic decisions and approaches used in my art practice.

3.1. Exploratory Studies

In addition to the two collaborative projects presented in this thesis, in chapters 4 and 5, a further line of work that relates to this thesis has been produced between 2013 and 2015. The exploratory studies consist of two projects that I have produced as audiovisual practitioner, which are part of the method used in this research.

The development of these projects allowed me to explore and experiment solutions that were later incorporated in this research project. The lessons learned during their development have also assisted me in improving my technical skills as a programmer and audiovisual practitioner. The solutions found throughout the development of these exploratory studies were later refined and incorporated into the practical portion of this research.

The first project, entitled *LiveRoom*, consists of an interactive virtual music listening environment. The application processes audio and video information simultaneously so as to create an audiovisual environment. Therefore, it provides this research project with valuable information regarding audiovisual composition within 3D space. The second exploratory study, entitled *12 Tones*, consists of a real-time sound visualisation application for live performance that contributes to this research project with an approach for applying audio data to control visual content.

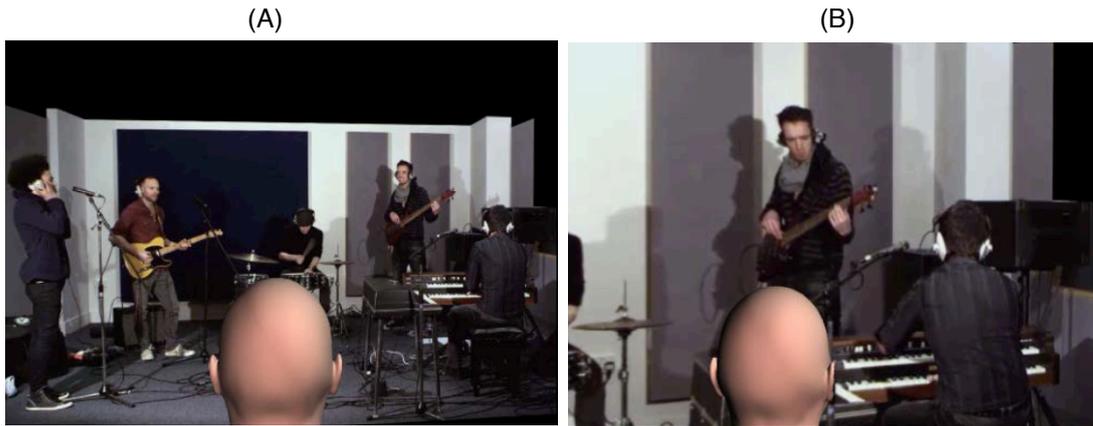


Figure 3.3. LiveRoom Music Listening Environment
 (A) Initial Position
 (B) Zooming in the keyboards

The project was developed and implemented in Pure Data programming language (Puckette, 1996), using Pd-Extended software (Puckette, 1996), where all the individual audio stems, videos, and 3D objects were introduced, and the audio and visual processing and interactive system were implemented. LiveRoom was presented at the 134th Audio Engineering Society Convention in Rome, in May 2013 (Figure 3.4), and received the Gold Award for Best Design Project during the Student Design Competition (Appendix A).



Figure 3.4. Presenting LiveRoom at the 134th Audio Engineering Society Convention in Rome, in May 2013.

LiveRoom places the participants in the centre of a virtual room, enabling them to act as producers by navigating within 3D space and immediately changing the audible response (Figure 3.5). The goal of the project is to convey acoustical messages from the sound sources to the listener, as it would happen in a real-world scenario. For the purpose of this project, audio and video footage of three songs performed by the band Citizen

(2013) were captured. In order to achieve accurate results, several stereo microphone techniques and close microphone techniques were employed during the recording process, and, ultimately, combined in the mixing stage.

As the listener navigates through the immersive listening environment, the distinct sound sources are subjected to several changes in level, balance, reverberation, panning and equalisation. Accordingly, it artificially re-creates the listener's ability to identify the localisation and origin of a produced sound.

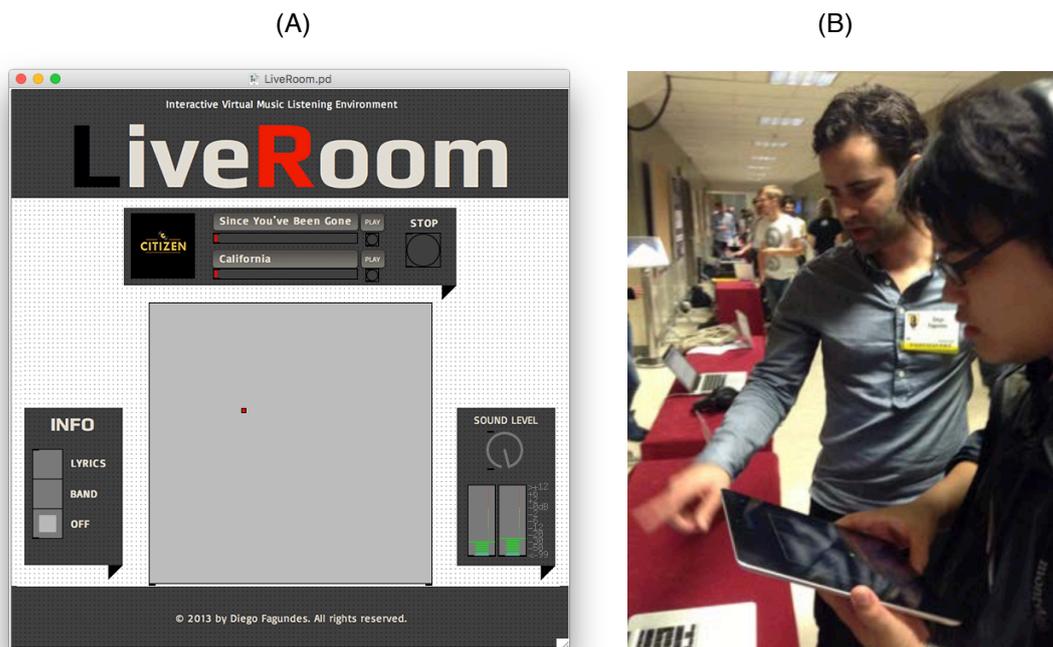


Figure 3.5. (A) LiveRoom Graphical User interface.
(B) User Interacting with LiveRoom at the 134th AES Convention, Rome, Italy, May 2013.

This project considers vital psychoacoustic factors in its design, such as the “audiovisual contract” (Chion, 1994: 115) and “auditory scene analysis” (Bregman, 1999: 641). The listener position is represented by a 3D model of a dummy head, which can be rotated and moved inside the 3D virtual room in real-time, generating an audiovisual relationship between the visual performance and the audible response (Figure 3.6).

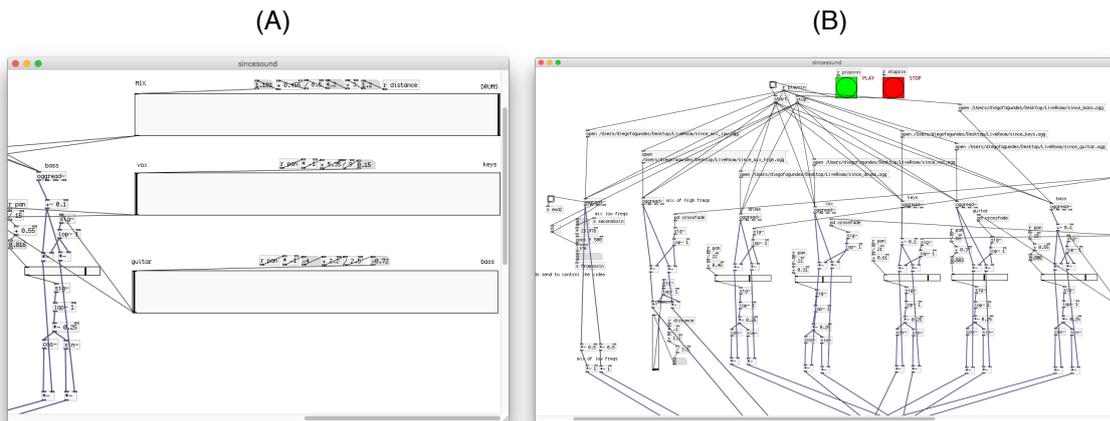


Figure 3.6. (A) The crossfades control the audio level between the instruments based on the 3D dummy head placement within the virtual room.

(B) Multiple audio players provide real-time processing for all musical instruments individually.

For instance, by placing the dummy head closer to one of the musicians, the user can visually and sonically focus on a specific musical instrument, changing its level, reverberation, panning and frequency response. Simultaneously, all the other instruments have their audio response also affected, interpreting the listener's perspective in relation to the sound sources.

The application contains a series of interconnected signal processing approaches that emulate fundamental elements associated with human hearing. As an example, a number of crossfades and Low/High pass filters recreate the interaural intensity difference by reconstructing in real-time a variable difference in loudness between the two ears. This effect is controlled by zoom and rotation movements, which are generated by the participants controlling the three-dimensional dummy-head object. Reverberation and other time-based effects are employed to emulate the interaural time difference, which consists of the difference in time that sound takes to reach both ears (Strutt, 1907: 214).

The user interface was designed to function as the platform for interaction. The grid displayed in the centre allows the user to control all the audio and visual processing. Although the project required the programming

of a complex audio processing system in order to accomplish these particular results, the participants' control is very intuitive.

3.1.2. 12 Tones - Real-time Sound Visualisation for Live Performance



Figure 3.7. 12 Tones Visual Display (Fagundes, 2014).

12 Tones consists of an application that analyses particular properties of audio signals and applies it to the generation and direct control of visual content (Figure 3.7). The system's architecture, as described in Appendix B, is directly related to the development of this research as it demonstrates a new approach for generating and controlling dynamic visuals using audio data. This concept was later further explored, refined and applied to the creation of the practical portion of this research.

This small study combines real-time processing and interactivity into a single tool, which is designed to generate multimedia displays in live performance scenarios. *12 Tones* was presented at the 137th Audio Engineering Society Convention in Los Angeles in October 2014 (Figure 3.8), and received the Silver Award during the Student Design Competition (Appendix B).



Figure 3.8. Presenting 12 Tones at the 137th Audio Engineering Society Convention in Los Angeles, USA, in October 2014.

Sound visualisation applications can be grouped into two distinct categories. The first comprises image sequencers, which provide playback, transitions and effects to moving images previously produced. The second category encompasses image generators, which support the creation of graphics, usually generated to complement musical material (Collopy et al., 1999: 111).

This project integrates both approaches, enabling the performer to generate graphics in real-time, as well as to control the movement and behaviour of images previously created. 12 Tones generates visual material using information extracted from the audio input signal's amplitude, tempo and pitch. Amplitude and tempo information are used to control the graphics' scale, intensity, transitions and rate; while pitch information is used to design a system where each of the twelve musical notes, when played, is responsible for triggering a particular colour (Figure 3.9).

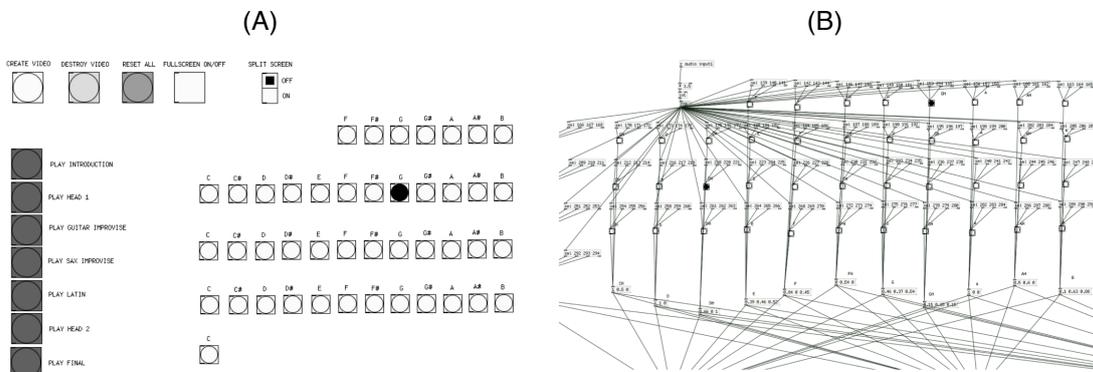


Figure 3.9. (A) 12 Tones Graphical User Interface
(B) Twelve sets of frequency ranges trigger each of the 12 colours.

The system was developed and embedded in the visual programming language Pure Data, which supports the combination and control of the audiovisual material (Puckette, 1996). 12 Tones uses information extracted from the audio input for two purposes: 1) generating visual graphics, and 2) controlling transitions and effects. The algorithm detects changes in frequency and sound envelope (Attack, Decay, Sustain and Release). As a result, it is able to output geometric shapes with different, colours, shapes, and sizes in synchrony with the audio input.

The most important aspect of the audiovisual system lies in the interaction between the performer and the system. This project uses the properties of the audio input signal produced by the performer, and uses it to generate and control the graphics. Therefore, it creates a synchronised activity between the audio and visual material.

Chapter 4. Case Study 1 - Surround Sounds

“It may seem like a contradiction, but I am interested in sound, not just for how it sounds, but also for how it looks”.

(Christian Marclay, 2014)

4.1. Introduction

Surround Sounds consists of large-scale multiscreen animation, presented by visual artist Christian Marclay (2015), in which I have worked as animator and audiovisual designer for a period of eight months. The project was presented during Christian Marclay’s solo exhibition at the White Cube art gallery in London, in January 2015. The project is a room-sized immersive video installation, in which onomatopoeic words, taken from the pages of vintage comic books, were animated and projected across four walls of a large room, from floor to ceiling. The central idea of the project is to translate *sound* into *motion*. That is, thousands of onomatopoeias were animated in such a ways that their movements would resemble the way they would sound.

This chapter explains the techniques I have applied to animate onomatopoeic images and to make them behave and perform, in terms of motion, in a manner that resembles or evokes the sounds they suggest. The development of this collaborative projects plays an important role in my research output since it allowed me to apply a technique for extracting information from audio files and apply it to the visual domain so as to produce visual motion that was later refined and applied to the development of the practical portion of this research.

My role in the collaboration process was twofold. Firstly, I was responsible for materialising and implementing, in practical terms, the concepts of motion that Christian Marclay had for each group of sounds that were previously selected and arranged. Secondly, I was required to create variations and new styles of motions that resembled the acoustic proprieties

The section 4.2 of this chapter outlines the list of sounds used to animate the onomatopoeias and demonstrates how they were organised. From this library of moving onomatopoeias, Surround Sounds was finally assembled. The final piece was a twelve minutes narrative orchestrated by Marclay as one large video section, with the dimension of 7690p x 1080p, which was designed to play as an endless and seamless loop, fully covering four walls of a dark room.

Surround Sounds is an artwork that challenges silence and encourages the audience to imagine sound itself. The essential purpose of the art installation is to produce an environment where the audience can experience a similar sensation to what they might experience if they walked into an intense and loud graphic novel, but while the installation remains entirely silent. Surrounds Sounds communicates through individual's recollection of noise, and invites the audience to build an imaginary auditory narrative, which is both personal and collective. Marclay suggests that "all onomatopoeias possess a particular sound quality" (2015a). Obviously, they do not produce sound, but they have the ability to invoke it.

In Surround Sounds, the onomatopoeias act as cues that send the audience in the search for sounds; hence, it is up to the audience to generate the sounds in their minds. As a result, the combination of visual motion, imaginary sounds, vivid colours and bold expressive fonts proposes a series of actions within a storyline. For instance, in the Surround Sounds piece the onomatopoeias BAM and BOOM seem to explode outwards in colourful blasts, while SHHH and FHhh slide down resembling dreary rain. *KRAK* accurately cracks in pieces across the screen, and *ZRZRZRZR* bobbles about at the lower portion of the screen like and a noisy wasp confined in bottle.

The intention of Surround Sounds is to express sonic information in visual terms. Therefore, in order to put Marclay's vision into practice, I was responsible for creating one model of motion for each of the onomatopoeias previously selected.

Throughout the creative process, hundreds of onomatopoeias were animated. Initially, the animations were created with the use of keyframe

animation. Then, later during the process, I presented a new approach, which consisted of extracting data from audio files in order to generate motion. This thesis will focus on this particular technique, which consists of a series of creative steps that resulted, according to Marclay, to be “the most effective and most used animation method throughout the development of Surround Sounds” (2014c).

This technique consists of extracting data from audio files and converting it into visual motion. This approach involves analysing an audio signal, extracting data from audio files and applying it to the control of the image’s opacity, position, scale and rotation parameters. This case study is an important guide for the development of this research project as it outlines an effective technique which I have implemented as an audiovisual practitioner engaged in a collaborative process.

This chapter contributes, in practical terms, to the exploration of the relationship between visual and sonic information, which was initiated in chapter 2.5 of this thesis. The techniques presented in this chapter, which include: extracting data from audio files, the use of sound data to generate motion, programming motion, programming randomness, were applied to the development of the practical portion of this research in order to generate visual motion.

Although this chapter illustrates the techniques employed using After Effects and its script programming language, the concepts presented in this chapter can be applied in many contexts. That is, they can be applied in any nonlinear editing, 3D animation, or motion graphic software, and the scripts can be translated and applied to any programming language.

4.2. Project Development Process

In order to create Surround sounds, Marclay assembled a huge collection of comic books, cut out all the onomatopoeias he could find on them, and classified them according to their sonic qualities. In some cases, he cut the onomatopoeias around their letters and, in other cases, he maintained

the background behind them. The result was the creation of stacks of folders filled with onomatopoeias, arranged based on the nature of sounds they evoke, such as: watery sounds, explosions, cracks, clicks, windy sounds and so forth.

WATERY SOUNDS	Blooop / Drip / Flooosh / Glip / Phloosh / Shloorp / Skluuurshh Ploosh / Spalooosh / Splaash / Splact / Splakk / Splaffff / Splat / Spletch / Splish / Sploosh / Splunck / Squish	Water fountain Footsteps, Water Splashes Kitchen faucet Underwater rumble River flow Water Pour Bath fill
EXPLOSIONS	Foom / Fashoom / Faboom / Doom / Dak-Koom / Croom / Cracaboom / Choom / Chonnk / Chakoon / Bzap / Bwa-Whoom / Brrrom / Brooom / Brakka-Thom / Boom / Badoom / Badaam / Choom / Faboom / Kaboom / Pa-choom / Shaakoom / Thoom / Vadoom / Whoom	Fireworks distant Fireworks crackle Distant bombing Nuclear Explosion Tank Explosion Short Explosion Reverberating Explosion Retro SFX Explosion Small Rocket Water Explosion
GUNSHOTS	Bada Bada / Badow / Bam Bam / Bam Wham Thram / BamBam / Bamf / Bang Bang / Bbrraaapppp / Bdam / Bdoom Bdoom / Blaam / Blaam Blaam / Bloom / Braap / Bratata / Buddabudda / Bwam / Kablam / Ratatata / Ratta Tatta / Sbang / Whomp / Zbammo	Machine gun shots Fireworks War zone Heavy machine gun Pistol shots Synthetic gunshot Laser gun
WINDY SOUNDS	Chsss / Fffff / Fffttsss / Fsss / Fwish / Hssss / Pffss / Psshhh / Slish / Sliff / Srsh / Sssss / Swishh /	Wind through window Wind ambience Sea breeze Dark ambience Medium wind Strong wind Wind Storm
ELECTRIC SOUNDS	Briiing / Bzzzt / Click / Crckle / Flazzz / Friiizz / Frazt / Fwakov / Fazzzp / Fzaap / Fzaam / Fziim / Kkrzzakk / Klik / Krinnng / Kzzz / Riing / Shazzakt / Skrazzt / Tzzzt / Zaaat / Zaaasp / Zam / ZipZap / Zoot / Zoop / Zaakkk / Zaatgot / ZZZraap / Zzzz	Circular saw Sound of bees Synthesizer random notes Freezer sound Radio signal sounds Ambien space sound Robot sound
HIGH PITCHED	Creaaak / Creeee / Eeeee / Eeeuuu / Frweee / Hssss / Hhhhss / Keesh / Kree / Kssshh / Screeee / Rheeee / Scckkkrrreee / Shreeee / Spanng / Squeee / Vreeee	Whistle Female scream Guitar harmonics Synthesizer high pitch notes Scratching sound
SLAPPING/ PUNCHING	Boop / Clap / Fap / Kapow / Pow / Powpoww / Slap / Smack / Smat / Thap / Thwap / Thwap-clap	Mouse click Button click Slap real Guitar slap Punch and slap Snare drum dry

Table 4.1. The Table of Sounds displays the groups of sounds created by Marclay (2014c) and the corresponding audio files used to generate visual motion (Fagundes, 2014b)

The process of creating the movement animations for *Surround Sounds* went through numerous development iterations for the period of eight months, during which the animations of all the individual onomatopoeias have been constantly refined. When I started working on the project with Christian Marclay, his original idea was to create an artwork approaching it as “piece of music”. As Marclay (2015) advises: “*the onomatopoeias are sound first, but in Surround Sounds, they become storytelling cues*”. Thus, the development of the project followed the same creative process that a music composition process would require; playing with words so as to explore and discover their potential, sonic wise, and subsequently start composing a piece.

The first attempts to animate the onomatopoeias revealed that the motions that can be created to resemble sounds could be classified into two categories. Firstly, the group of moving onomatopoeias that seem to resemble the way objects resonate while they vibrate. Secondly, the group of moving onomatopoeias that seem to shout the sounds they represent. As a result, in order to generate motion for the first group of onomatopoeias, I extracted data from sound files and used it to create a library of motions (Table 4.1), that later could be applied to the onomatopoeias. The technique used was based on the Motion Math Script method, which was part of the early versions of Adobe After Effects until 2001 and the convert audio to key frames feature found in Adobe After Effect CS6.

I started by recording and collecting a series of sounds (Table 4.1) generated by objects being struck, plucked, hit and strummed. Next, I extracted the information from these audio files, using the technique described in the sections 4.3, 4.4 and 4.5 of this chapter, and directly applied it to the layers’ X, Y and Z parameters, so as to drive their animation. This approach is directly linked with the development of the algorithms presented in this research as it informs an approach of linking audio and visual information, as it aims to translate audio data into the visual domain creating a direct relationship between sonic and visual information.

4.3. Programming Motion

Surround Sounds was created using Adobe After Effects with a combination of keyframe animation and the implementation of bespoke JavaScript expressions. The use of expressions was particularly important for the creation of complex animations, as this approach reduces the need for manually created keyframe animation, whilst simplifying the process of applying specific motion behaviour to multiple images. The expression language is a JavaScript based programming language that permits the manipulation of time-based streams. The expression itself is a small piece of software, which calculates a given value for a specific layer at a specific moment in time, and can be used to create relationships between different objects (adobe.com, 2016). For instance, a specific property of one layer can be used to dynamically animate another layers with the use of a single line of code.

```
Code:  
// the amplitude of the audio file is used to control the position of a layer  
// this is a very simple one line script  
  
thisComp.layer("sound_amplitude").transform.position
```

Figure 4.2. A specific propriety of an audio file controls the position of a different layer.

The expressions in Adobe After Effects are written in standard JavaScript programming language inside the individual layers and operate directly on the property values. The expression takes the value of a property, according to what is determined by its keyframes, and utilises it as an input value to generate and modify their values. Therefore, it generates variations adding a more organic feel to the animations. For instance, the `wiggle()` expressions, which was continually used throughout this project, as outlined in detail in chapter 4.6, can be employed to randomly move an object within pre-defined distance ranges.

Since After Effects uses *Javascript* programming language, and the practical outcome of this research will be implemented using *Processing.js* to translate code written in *Java* into *Javascript*, both projects behave in a similar

manner programming wise. As a result, the techniques developed for Surround Sounds will be an important element for the development of the software that this research expects to deliver.

4.3.1. Extracting Data from Audio Files

For the purpose of illustration, this section outlines the process of animating the onomatopoeias that are part of the Surround Sounds piece, so as to fully demonstrate how the technique of extracting data from audio files was created and implemented. As advised by Christian Marclay, the onomatopoeia THOOOM should visually behave, in terms of motion, in a manner that would resemble the sound of a loud alternating stroke of a drumbeat. To accomplish this task, I began by recording a synthesised drum sound, being struck for a period of ten seconds. The resulting audio file consisted of a close captured drum being continuously hit at 120bpm with a sound envelope that contained a fast attack, short decay, short release, short sustain and an amplitude of -6 dBFS. The next step was converting the audio frequency-time domain data into a visual graphic representation showing the varying audio amplitude over time, formed by hundreds of keyframes, which was achieved with the use of the keyframe assistant function (Figure 4.3).

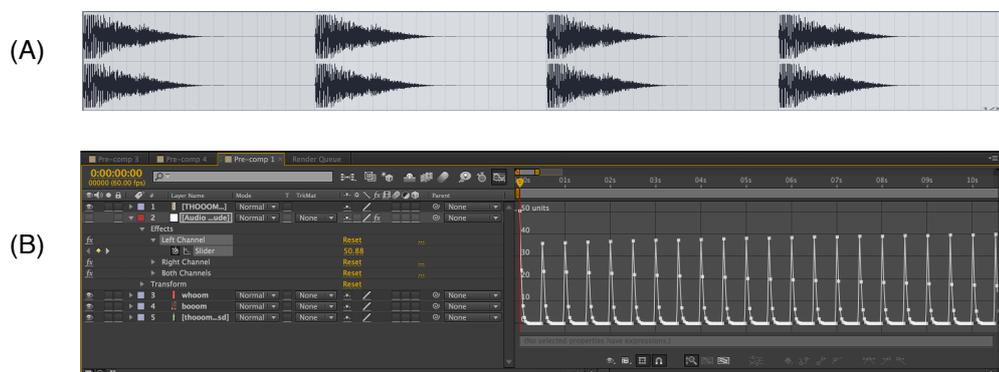


Figure 4.3. Converting audio into keyframes

(A) Drum sound recording waveform

(B) Audio signal represented as a series of keyframes.

The graph generated from the audio data does not differentiate frequencies; that is to say, it only visualises the amplitudes. For that reason, in order to achieve a more accurate representation of the range of frequency desired within the audio spectrum, an audio equaliser was used to boost or attenuate certain frequencies before converting the audio files into visual graphs. For instance, the audio file used to generate the motion of the onomatopoeia HONK, which consists of the recording of a car horn, had its mid frequency range, between 800Hz and 1200Hz, boosted in order to accentuate the frequency range that presents the higher concentration of energy so as to have a more accurate visual representation (Figure 4.4).

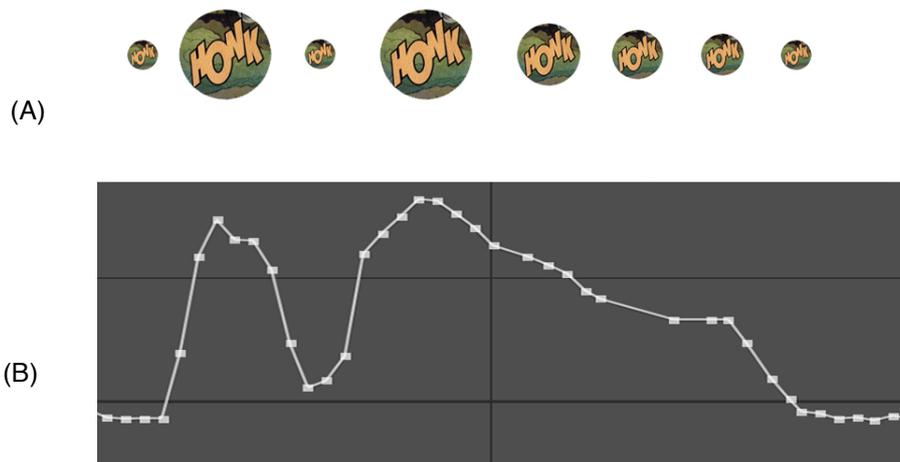


Figure 4.4. (A) Onomatopoeia scale being controlled by the audio amplitude.
 (B) Car Horn audio file represented as keyframes

In the same fashion, image below (Figure 4.5) shows the data extracted from a machine-gun sound recording being applied to the opacity of the onomatopoeia RATTATATA. In the following example, the majority of transients are located in the high frequency range, above 5000Hz. Thus, a shelf audio equaliser was used to boost this range of frequencies. Unquestionably, the equalisation method was an important step included in this technique. That is, accentuating a specific range of frequency that more accurately represents the nature of a specific sound leads to a movement that better translate the ways an object would vibrate and sound.

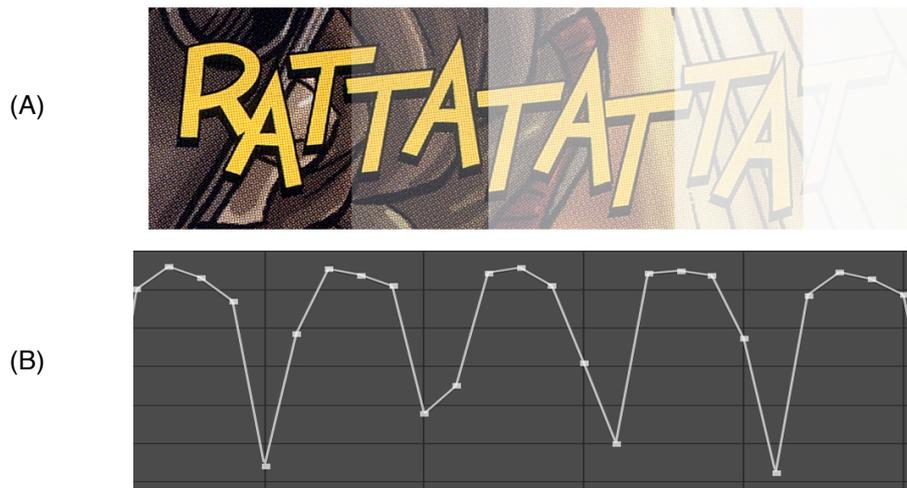


Figure 4.5. Using audio to control opacity
 (A) Image changing opacity
 (B) Machine gun audio as keyframes

It is important to mention that this animation process could also be achieved manually. However, it would be an extremely time consuming activity, because it would be crucial to identify each of the peaks in the audio amplitude and place keyframes, one at a time, matching each of the peaks. Hence, since this technique was going to be repeatedly used throughout the project, the creation of an automatic system was required to facilitate the workflow and to insure that more time would be assigned to the creative process, rather than to time consuming repetitive tasks.

4.4. Converting Audio Data into Motion

Once the data was successfully extracted from the audio files, it was necessary to convert this information into readable data, so it could be implemented to drive visual motion. The image below (Figure 4.5) shows the onomatopoeia THOOOM, which had its height scale controlled by data extracted from an audio file. The animation resulted in the image being vertically extended and shrunk in synchrony with the audio. As the image below illustrates (Figure 4.6), the highest peak of the waveform represents the point in which the image is extended to its normal ratio, while the lower portion

of the graph, in this case 0, represents the point where the image is shrunk at a proportion of 10%. Thus, the motion of the onomatopoeia THOOM represents a short burst sound with a fast attack and high amplitude transient and fast decay and release.

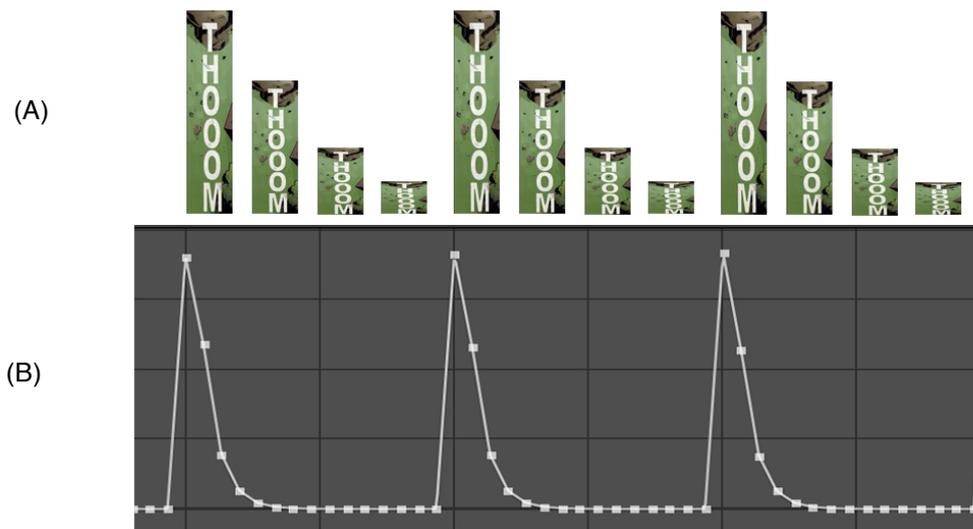


Figure 4.6. (A) Onomatopoeia Y-axis scale is controlled by the audio amplitude.

(B) Audio Waveform

As the code below shows, the expression used to generate motion specifies the audio's highest and lowest amplitude levels with `minAudioToom` and `maxAudioToom`. The variable `minAudioToom` represents the minimum value that the audio amplitude will reach, while the variable `maxAudioToom` denotes the maximum value that the audio amplitude will reach. The `maxStretchToom` represents the amount by which to multiply the layer's scale at maximum audio amplitude.

The next step was to attach the variable `audioLev` to the slider containing the audio amplitude keyframes, then, to use the `linear()` expression so as to translate the current audio level to a suitable stretch value. The minimum stretch value was set to 1.0, which will leave the expression scale value unchanged. Finally, the expression scale value was multiplied by the stretch factor.

Code:

```
// the amplitude of an audio files is being used to control the vertical scale of an image
// at maximum value image will be stretched
// at minimum value image will be shrink
// setting maximum and minimum scale values

minAudioToom = 0;
maxAudioToom = 15;
maxStretchToom = 2.0;

audioLev = thisComp.layer("Toom Amplitude").effect("Both Channels")("Slider");
stretch = linear (audioLev, minAudioToom, maxAudioToom, 1.0, maxStretchToom);
value * stretch
```

Figure 4.7. The height value of a layer is being controlled by the data extracted from an audio file.

As the code above demonstrates, the `linear()` expression was included in order to apply the information extracted from the audio file into the height scale parameter of the layer. The `linear()` expression consists of a linear conversion from one range of numbers to another. It returns a value that linearly interpolates between two values as the input varies. The `linear()` expression is, in this case, responsible for tying together the changing values expressed as keyframes to the image height scale parameter. As a result, the animation is directly matched to the visual graph that is graphically representing the audio file.

4.5. Creating Motion Randomness

As has been previously stated, the aim of Surround Sounds was to generate motions that resemble sounds. Therefore, an important aspect to be taken into account is the apparent randomness present in real world acoustic environments, which are produced by the reflections originated from the complexity of surfaces and materials found in physical spaces (Kelly et al., 2016).

Thus, in order to add this characteristic that is naturally perceived in the realm of sound, I implemented a method using the `wiggle()` expression. The `wiggle()` expression is a method for randomly modifying a property value. As

opposed to the predicted positions generated by keyframe animation, which is done manually, the `wiggle()` expression always generate random positions to all parameters it is applied to, by randomly shaking that specific value.

The `wiggle()` method can take up to five parameters. However, for the purposes of this project, only two parameters were taken into account. The first, frequency, which defines the number of wiggles per second, and the second, amplitude, which defines in its unities the property being wiggled. For example, if the property being wiggled is position, the amplitude will comprise pixels, if the property being wiggle is rotation, the amplitude will represents degrees

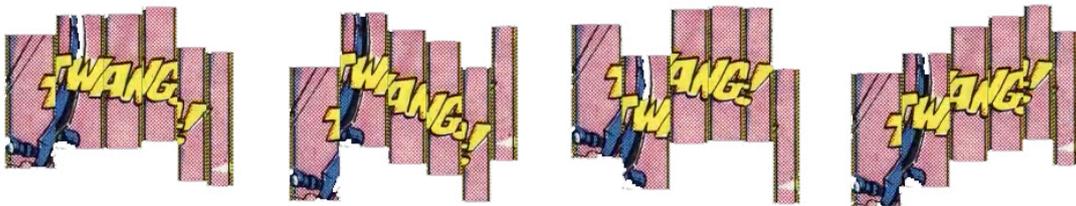


Figure 4.8. The wiggle method accentuating the Y-axis movement

The image above, display an example of the onomatopoeia TWANG (Figure 4.6). This onomatopoeia was sliced in numerous vertical pieces and each of these individual pieces had the `wiggle()` method applied to it containing discrete values.

Code:

```
wiggle(3,15);
```

Figure 4.9. The wiggle expression.

The `wiggle()` method can be applied to both X and Y axis independently. Nevertheless, in the case of the TSSSS onomatopoeia the expression is requesting the layer to wiggle the X-axis property of the layer three times per second, between the values of -15 and 15 pixels.

4.6. Technical evaluation

The main goal of this collaborative project, that of translating sound into motion, was successfully achieved. From an audiovisual designer's perspective, I believe that the technique implemented to extract data from audio files and directly apply them to animate parameters of images added an organic and realistic feel to the final piece in a very effective way. Therefore, it directly influenced the way in which the audiovisual art installation was created and perceived.

In the final analysis, it has been noted by Marclay as well as all members of the development team that "the motion created from the data extracted from audio files presented a more natural feel, compared to the ones that were manually animated with keyframes" (Marclay, 2014). To explain, the motion created with the technique presented in this chapter, closely resemble the way in which objects vibrate and resonate to specific sonic impulses. Thus, because it emulates phenomena present in the real world, the technique of using data extracted from audio files to generate motion presented a more organic feel when compared to the motion created using traditional keyframe animation. The approach used to organise the onomatopoeias into groups based on their sound quality, was essential to support the creation of the final piece, because it accelerated the process of assembling the moving onomatopoeias as a musical piece, where all events were connected to each other. For instance, a series of onomatopoeias that move in such a way that resemble a humbling low frequency noise are followed by an explosion, which in turn is followed by fragments of onomatopoeias that fly down the screen as leaves. Ultimately, Surround Sounds plays like a musical composition, displaying visual material that suggests aural information; a silent symphony of illustrated words that plays through our recollection of sounds. Since this approach focuses on specific frequency ranges for each group of sounds, an automated version of this technique could directly assign those frequencies to specific parameters of an

image. This automated tools is expected to be developed and presented as part of the practical outcome of this thesis.

4.7. Impact of Surround Sounds

Surround Sounds was presented during Christian Marclay's solo exhibition at the White Cube gallery in London, in January 2015. As previously stated, Marclay intended to create a visual artwork that could be read, or mentally played as a music score. Surround Sounds proposes a private mode of listening. In other words, it stimulates a complex and layered mental performance, which invokes diverse classes of sounds and genres of music that, according to Marclay (2015), are unreachable by means of sound recording and music performance.

Surround Sounds explores the embedded ambiguity of sound. Thus, it suggests that it is impossible to think of mute images, in the same way that it is impossible to imagine a soundless imagination. As Marclay explains in an interview for the newspaper the New York Times (newyorktimes.com, 2015):

“I want it to be aggressive, I want it to be physical. I think you have to walk out of there feeling kind of a bit uneasy and dizzy.”

Despite being silent, Surround Sounds draws from music and sonic culture in general, and without the actual manifestation of sounds, Marclay aims to create an interchange between an audible visual world and a muted sonic domain. Surround Sounds is a composition that turns each viewer into a composer, as well as performer of mental music. Therefore, the four walls, which are filled with moving onomatopoeias, became restructured into an enormous score, as the room itself becomes a real stage for conceptual audiovisual improvisation.

The outcome of this collaborative effort is central for the development of this thesis, in both conceptual and practical terms as the techniques created during its development were further developed, in a later stage, and implemented in the practical portion of this thesis. Also, the approaches with

which Surround Sounds explores the relationship between sound and image are also important guides to inform the development of this research project. My work as audiovisual designer in this collaborative project with Christian Marclay has been very fruitful. In particular, because it endorsed the creation of the techniques outlined in this chapter.

Chapter 5. Case Study 2 - Interactive Art Gallery

“The artist builds models of the new environments and new social lives that are the hidden potential of new technology”

(McLuhan, 1964b: 49)

5.1. Introduction

This chapter describes the second case study presented in this thesis, which consists of a framework for combining visual artworks, music, sound design and non-linear narratives for the creation of screen-based aesthetic experiences, a process called *Interactive Art Gallery*. The platform was developed through a process of iterative design in collaboration with two visual artists represented by the Debut Contemporary Gallery, during which a series of three screen-based artworks were produced. The project was presented at the 136th Audio Engineering Society Convention in Berlin, in April 2014, and received the *Gold Award for Best Student Design Project* (Appendix C).

The development of Interactive Art Gallery includes the use of a set of specific techniques as follows: image segmentation; texture synthesis (inpainting); 3D composition; application of depth of field blur; 3D lighting and virtual camera animation; music composition; and narrative design. The central idea of this framework is to extract and manipulate elements of an image to reflect their content within 3D space. Additionally, crucial to the frame, is the creation of an interactive system to allow the viewer to navigate through a non-linear narrative, which is presented in the form of separate *loops* that can be played in any given order. Interactive Art Gallery was implemented using Adobe Flash as it seemed to be a suitable solution for embedding the project in the Debut Contemporary Website. As the diagram below (Figure 5.1) reveals, the separate loops comprise a scene of the narrative, which independently display features related to the artworks. As result, the sum of the loops, which is represented by the overlapping section of the diagram, provides the meaning of the overall narrative.

This chapter outlines what are the essential techniques and approaches I used to develop the series of screen-based artworks. Then, it outlines the design principle and methods used to generate them. Therefore, this chapter assists this research project by identifying what the central opportunities are for developing automatic systems that can perform, in a more efficient way, the main tasks undertaken during the creative process. The implementation of these techniques as algorithms is later described during the development of the Audiovisual Collage software, in chapter 6 of this thesis.

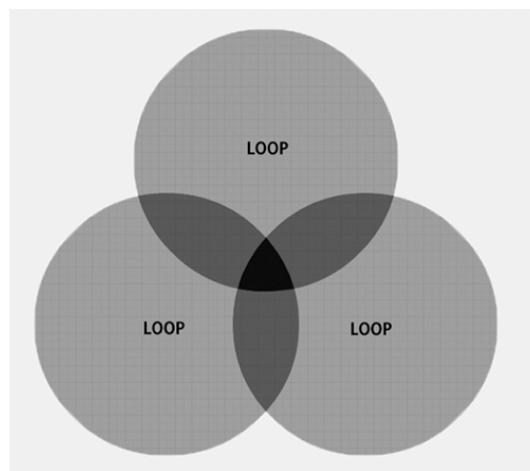


Figure 5.1. The overlapping of loops creates *meaning* in the IAG; where each loop comprises a particular scene of the narrative (Fagundes, 2014).

The aim of this platform is to engage viewers with narratives that present technical aspects and artistic qualities of the works of art presented, through the perspectives of the artists themselves. As a result, this creative process reveals new conventions that can be created to take advantage of the affordances of new delivery formats within the digital medium. Furthermore, this chapter explores how the application of a digital set of techniques and tools can be used to serve multiple forms of art. The power of this project lies in its generality as it can be applied to a wide range of visual art forms that include painting, sculpture, photography, film and documentary.

For this project, three screen-based interactive artworks were created in the context of a series of design collaboration, between June and September 2014, with three visual artists represented by the Debut

Contemporary art gallery: Carne Griffiths, Tahnee Lonsdale and T. Raymonzrek. The project was anticipated to promote and bring attention to these three particular artists who have their artworks commercialised by the gallery. The complete screen-based interactive series was on display during 2014 and 2015 in the Debut Contemporary website. The series can no longer be found online; nevertheless, they can be accessed in the accompanying folder [Appendix E]. The Interactive Art Gallery framework, which was developed as a precursor to the final AudioVisual Collage framework presented in this thesis, constitutes a case study for the analysis and evaluation of an on-going iterative and creative design process. Thus, the projects serve as an instrument for the exploration of potential technologies in the elaboration of future forms of audiovisual aesthetic experiences (Figure 5.2).

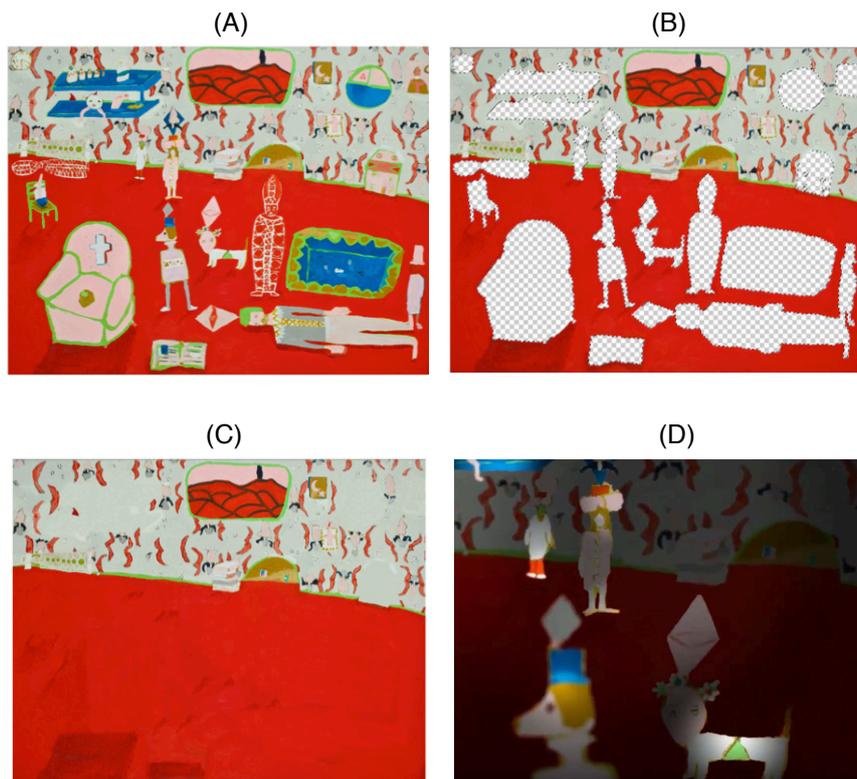


Figure 5.2. Interactive Art Gallery Development Stages:
(A) Original Painting *Vatican Slumber Party* (2014) by Tahnee Lonsdale
(B) Image Segmentation
(C) Inpainting
(D) 3D Composition, Depth of field blur and 3D Lights implementation

5.2. Project Development Process

The process began with open-ended discussions with the three artists participating in the collaborative process. The first topic was with regards to the direction and role that narrative would play in relation to their work, followed by the choice of the artworks to be explored. The pieces were particularly chosen based on their capacity to represent their creators' artistic technique and style. That is, the piece was chosen based on their capacity to convey their artists approach to art.

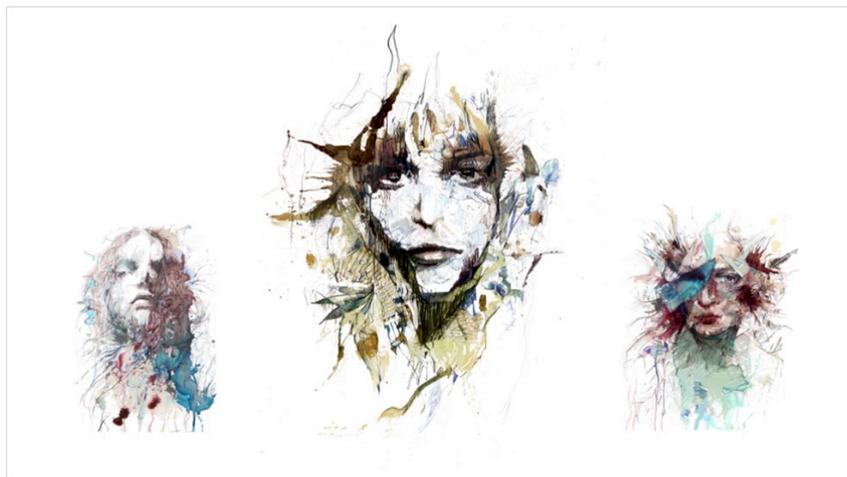


Figure 5.3. Painting *Another Place* by Carne Griffiths (2014).

The first piece chosen was *Another Place* (2014) by artist Carne Griffiths (Figure 5.3). According to the author, the chosen piece is: “the best representation of my approach to art on a single canvas” (Griffiths, 2014). The piece *Another Place* was chosen because we wanted to generate an audiovisual aesthetic experience that could potentially outline not only the individual work of art, but also the overall approaches used to create it.

The second piece entitled *Vatican Slumber Party* (2014) by artist Tahnee Lonsdale (Figure 5.4), was chosen because according Lonsdale, the piece can be seen as a strong representation of her technique and approach to art, on her words: “*Vatican Slumber Party* consists of a semi abstract,

staged scene, where the figures both hide and flaunt themselves to the viewer; leading them across the canvas (Lonsdale, 2014).



Figure 5.4. Painting *Vatican Slumber Party* by Tahnee Lonsdale (2014).

This chapter presents the evolving creative process of Interactive Art Gallery and identifies the development steps that could be automated with the creation of algorithms, which can assist other designers and artists to create their own screen based aesthetic experiences. Moreover, this chapter explores methods so as to find the ones that are more useful to indicate which one to implement later.

The following sections are detailed descriptions of the visual design principles and methods used in the development of Interactive Art Gallery. For each stage of the process, additional automated approaches to complete the same specific tasks are also presented. These additional approaches involve algorithms that require minimal user interaction to perform each of the particular assignments. Therefore, this chapter offers new directions that can be explored for the combination of existing algorithms approaches in the creation of original software, which will ultimately be implemented in the practical portion of this research.

5.3. Image Segmentation

The development of Interactive Art Gallery started with the extraction of the expressive elements contained in the artworks. For instance, during the exploration of the artwork *Vatican Slumber Party* (2014) by artist Tahnee Lonsdale, the artist was asked to pinpoint the most important elements of the artwork in relation to the narrative she wanted to create in the artwork. As a result, as the image below outlines (Figure 5.5), the elements identified by the artists were extracted from the original image.



Figure 5.5. (A) Painting *Vatican Slumber Party* (2014) by Tahnee Lonsdale
(B) Elements chosen by the artists to be extracted from the painting

There is not a unique approach to this topic, but several techniques that vary according to the nature of the elements to be extracted. This section presents the two techniques that were used to segment images during the development of Interactive Art Gallery, which consist of the *Quick Selection Tool* and the *Magnetic Lasso Tool* (Adobe.com, 2016), both found in Adobe Photoshop.

The *Quick Selection Tool* selects the pixels of the image based on their colour and tone. It works like a brush, allowing the user to select areas simply by painting over them. Nevertheless, when dealing with complex shapes the Quick Selection Tool often produces edges that need to be refined. In that case, the *Magnetic Lasso Tool*, which works as an edge detection tool, is a suitable alternative. The Magnetic Lasso Tool attempts to find the edges of an

object by searching for differences in colour and brightness values between the object and its background; then, it snaps the selection outline to the edge and attaches to it like a magnet. Ultimately, the extraction of the objects from the image's background is accomplished with the creation of new layers containing only the objects selected.

There have been several approaches to image segmentation. However, this research focuses on two particular methods that use only a small amount of interaction to perform the task of image segmentation.

The first approach, which was proposed by Boykov and Jolly (2001), is called *Interactive Graph Cuts*. It consists of a technique for interactive image segmentation that requires the user to identify certain areas as *foreground* or *background* so as to provide hard constraints for segmentation. In other words, it is an approach in which the user specifies the objects to be extracted as well as the final composition. The obtained solution gives a balance of boundary and region properties among all segmentations. However, the topology of segmentation is unrestricted, as both *object* and *background* segments may consist of several isolated parts. That is, it might sometimes present extracted regions that are not consistent and objects present in the image might have parts of it as foreground and other parts as background.

The second approach, which was proposed by Rother et al. (2004), is called *GrabCut*. It is a method for image segmentation that prompts the user to mark a single rectangle around the object to be extracted from the image. The outer part of the rectangle is defined as background, while its inner part as an unspecified combination of the object's foreground and some background. These constraints are used as an initial solution to the image segmentation problem, leading to an iterative method, which in conclusion aims to assign each pixel of the image as background or as foreground.

Those additional methods serve as important guides for the development of this research, as they offer interesting solutions for automating the image segmentation task. These techniques are further examined and defined in the development of Audiovisual Collage image segmentation algorithm.

5.4. Inpainting

After successfully segmenting the images, the next step of the creative process was to fill-in the gap in the background image, which was generated by the segmentation process, in a visually plausible way. The inpainting method used in the development of Interactive Art Gallery was the *Clone Brush Tool*, found in Adobe Photoshop (Adobe.com, 2016), which allows interactive painting using pixels from the original image (Figure 5.6).



Figure 5.6. (A) Gaps left in the original image
(B) Inpainted background image

The Clone Brush Tools is a well know approach to inpainting as it consist of a method for inpainting that present very good results, however, requires a high level of user interaction to be completed. The inpainting process is not simple to be achieved algorithmically (Wang and Liu, 2009), thus, this section explores methods so as to inform how to develop the automatic inpainting to be implemented in the practical portion of this thesis.

Similarly to automatic image segmentation, automatic inpainting is not generally available in tools such as Adobe Photoshop. Thus, this research evaluates different solution for the inpainting problem, so as to create an algorithm to automatic perform this task. The following automatic inpainting algorithm was implemented in an early prototype so as to perform a user

study, in order to ask other artists and designers if they would need and want such a tool to perform inpainting automatically.

The technique utilised in the development of Interactive Art Gallery was the Clone Brush Tool in Adobe Photoshop, which performs the duplication of certain portions of the image using a source, a destination, and a brush (Figure 5.7). The process involves choosing a sampling region that shares similar texture with the area that needs to be filled, and using it as a reference point to the new cloned area. That is to say, the clone tool translates the reference point into the region that needs to be covered. Even though pixels from the same image are more likely to be self-similar, the act of painting over the holes on the background image faces three challenges: first, choosing the right constant offset to the source pixels; second, choosing an appropriate brush size; and third, finding source pixels with structure that is compatible with the destination.

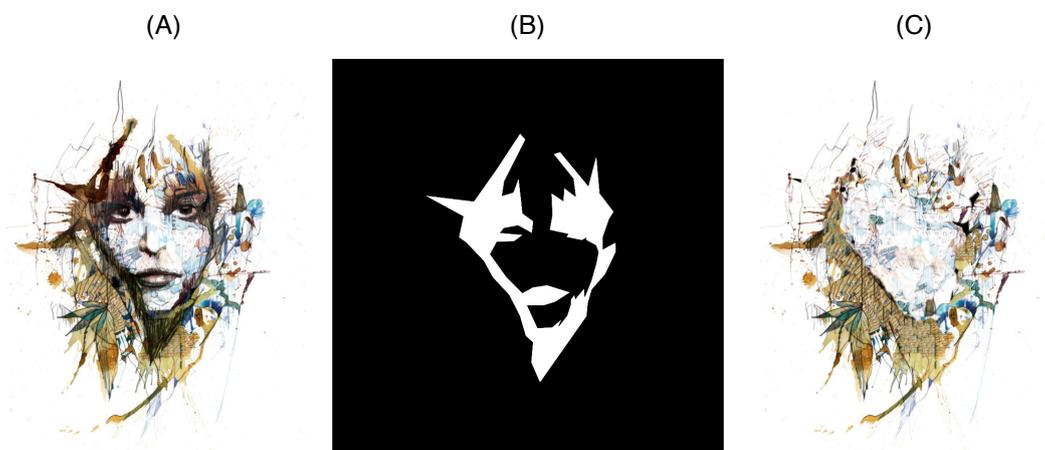


Figure 5.7. (A) Original image
(B) Area Extracted in Image segmentation process
(C) New inpainted background image

In the past, two classes of algorithms for the automatic repainting of background images were proposed: *texture synthesis* (Wei et al., 2009) that consists of algorithms for generating large image regions from sample textures; and *inpainting*, which according to research by Goyal and Diwakar (2010) is the process of reconstructing lost or deteriorated parts of images.

Nonetheless, in spite of being used in the main to remove images' imperfections, the inpainting process is also applied to the segmentation of objects.

This research project initially considered both texture synthesis and Inpainting approaches. However, it finally focuses on a specific method called *Exemplar-based Image Inpainting*, which was proposed by Criminisi, Perez and Toyama (2004). This method aims to automate the *clone tool* process by filling the missing areas of an image with similar patches located in nearby regions. To put simply, the algorithm performs the search and copies the pixels from the similar area into the gap. The main advantage of this approach is that it performs inpainting at a patch level, instead of pixel level. Therefore, this method lessens the amount of distorting and blurring artefacts generated by other methods. This approach is further analysed and implemented during the development of the inpainting algorithm, which is part of the practical outcome of this research.

5.5. 3D Composition

One of the key features of the Interactive Art Gallery is the creation of 3D movements from still images. This section discusses the techniques applied to accomplish this effect, which include: 3D Composition (the placement of images in 3D space); the implementation of 3D virtual camera, depth of field blur, and 3D lighting effect. Moreover, this section will outline some of the important aspects of *perspective* and *parallax*, and their aesthetic role in the Interactive Art Gallery project. In addition, this section describes these key elements with reference to how they can be achieved in traditional software, and also through the use of the Processing programming language, in order to establish a clear mechanism for realising automated methods for reproducing the specific creative workflow in question.

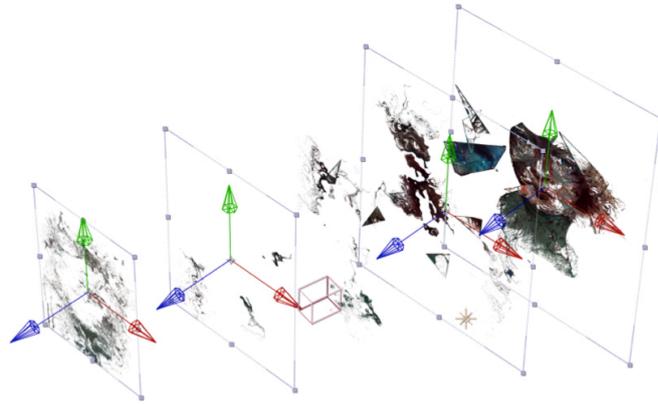


Figure 5.8. The Multiple layers of the piece Another Place placed within 3D Space

Once the processes of image segmentation and inpainting were completed, the images were split into multiple layers. The next stage of the process was to place these discrete layers within 3D space, where the 3D movements would later be created (Figure 5.8). The 3D movements were generated using tools found in Adobe After Effects (Adobe.com, 2016), which replicate analogue cinematic techniques (Figure 5.9).

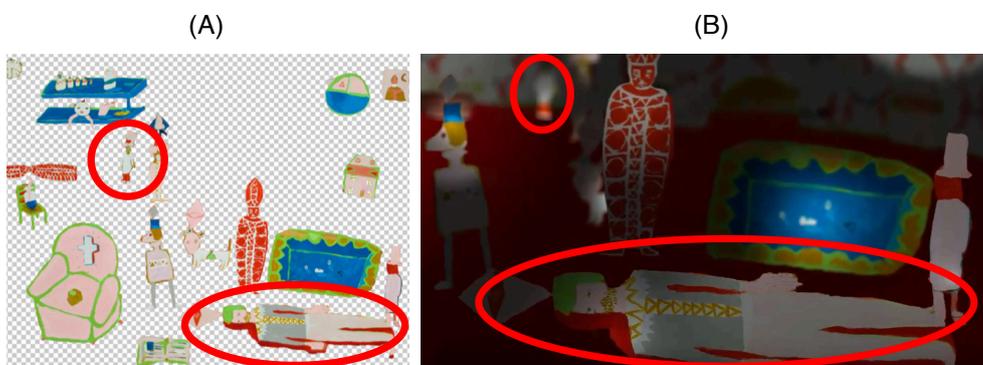


Figure 5.9. (A) Objects' original size
(B) Objects' relative size changed by perspective

A consequence of positioning the layers within 3D space is an effect called *parallax*, a natural result of 3D perspective. For instance, the images located closer to the viewer move faster than the images situated further away. Likewise, when an object flies towards the viewer, it will eventually go over their head. The more the objects are spread apart in space, the more the parallax effect will be noticeable (Kaneko et al., 2001).

To date, numerous methods have been exploited to position images within 3D space, including graphic editors, motion graphics software, and visual compositing applications. This project focuses on a method in particular, which can be implemented using the Processing programming language.

5.5.1. Virtual Camera

The 3D virtual camera was implemented using Adobe After Effects virtual camera tracking feature, which aims to faithfully emulate a real camera's structure and motion (Figure 5.10). Furthermore, according to Sulonen (2011), scenes captured through lenses within the range of 35mm to 90mm, generally contain about as much imagery as a person can perceive without moving her head. Therefore, the relationship between foreground and background objects appears correct and natural, and objects do not suffer from apparent distortion, except for the usual distortion generated from perspective.

The *camera depth of field blur* is a feature applied to 3D layers that allow them to be selectively sharpened or blurred out of focus, depending on their distance from the 3D camera. It consists of a faithful simulation of a real camera (Figure 5.10).



Figure 5.10. Camera Depth of Field effect in the piece *Another Place*

Processing offers a highly efficient automatic approach to virtual camera. When looking at a 3D scene through Processing window, one can think of the observer's view of the scene as a camera. In fact, there is no actual camera; this is just a convenient device to help one to understand how to traverse a 3D scene. The simulation of a camera's behaviour, in terms of movement, can be achieved through transformations inside `draw()` by using `translate()`, `rotate()`, and `scale()` to manipulate the observer's view of the scene. Still, for convenience there is also a `camera()` function whose purpose is also to simulate a camera. The function defines a camera as having an eye position, the camera location, a scene *centre*, which tells the camera which way to point, and an upward axis that aligns the camera vertically (Processing.org, 2016).

5.5.2. 3D Lights

Four types of 3D lights, which can be found in Adobe After Effects, were used in the development of Interactive Art Gallery. Firstly, the *Ambient Light*, which has no position or direction, as it illuminates all layers equally, no matter where they are located in a given scene. Secondly, the *Parallel Light*, that is supposed to represent the effect of a light located at a great distance, such as the sun, and have its light rays cast in the same direction, parallel to each other. Thirdly, the Spot Light, which emits light from a source that has a cone angle and a cone feather, which works like a flashlight or a spotlight utilised in stage productions. It was mainly used to follow objects being describes in the narrative. Finally, the *Point Light*, that functions as a light bulb hanging in space, which was mostly used to illuminate objects in dark scenes (Figure 5.11).

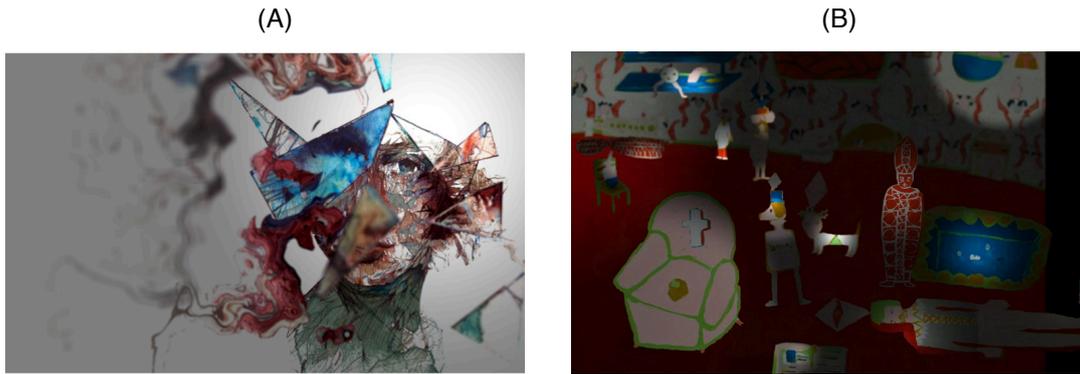


Figure 5.11. (A) Point light applied to the background layer in *Another Place*.
 (B) Multiple Point lights applied to individual elements of *Vatican Slumber Party*

Processing offers a solution for 3D lighting while running on P3D rendering mode, allowing the lighting of the 3D elements in the scene to be manipulated in real-time. In the same way that drawing in three-dimensions is an illusion, the addition of lighting to a Processing sketch is a simulation of the idea of real world lighting, which can be used for the purpose of creating a variety of effects. This effect is particularly useful because some objects, such as a sphere, do not appear three-dimensional until they are lit (Processing.org, 2016).

5.6. The Role of Sound in the Interactive Art Gallery

The sound track of Interactive Art Gallery was anticipated to reinforce the motion and texture of visual objects displayed on the screen. The motions observed in the audiovisual aesthetic experiences were generated with the employment of two distinct approaches. First, by animating a virtual camera in 3D space, which displays the observer's point of view. Second, by independently animating the elements extracted from the image, through image segmentation. Hence, the sound track is created as an attempt to synchronously match the motion of objects and sounds so as to create associations between the sonic and the visual textures of the objects. This case study project along with the first case study project, *Surround Sounds*,

consist of approaches to explore the relationship between sound and image and consist of important guides to inform the development of this research.

The image below (Figure 5.12) shows the artwork *Another Place* (Griffiths, 2014), which has a visual texture formed by glass fragments and liquids. As the camera flies through the fragments placed within 3D space, the sound tracks reinforce the visual texture, the shape and the movement of the objects, creating a relation between the auditory and visual elements. In this specific example, to created this relationship, the sound track was generated recorded watery and metallic sounds as well as high frequency synthesised noises and melodies (Table B).

Track 01	Dripping Water	Acoustic	Small Blobs of paint
Track 02	Shaking water bottle	Acoustic	Moving Liquids
Track 03	Water fountain	Acoustic	Large Blobs of paint
Track 04	Broken glass	Acoustic	Triangular Shapes
Track 05	Metallic ambience	Synthesised	Approaching face
Track 06	Feedback loop (Melody)	Synthesised	Moving Lines

Table 5.1. Table of Sounds recorded for the Interactive Art Gallery exploration of Another place.

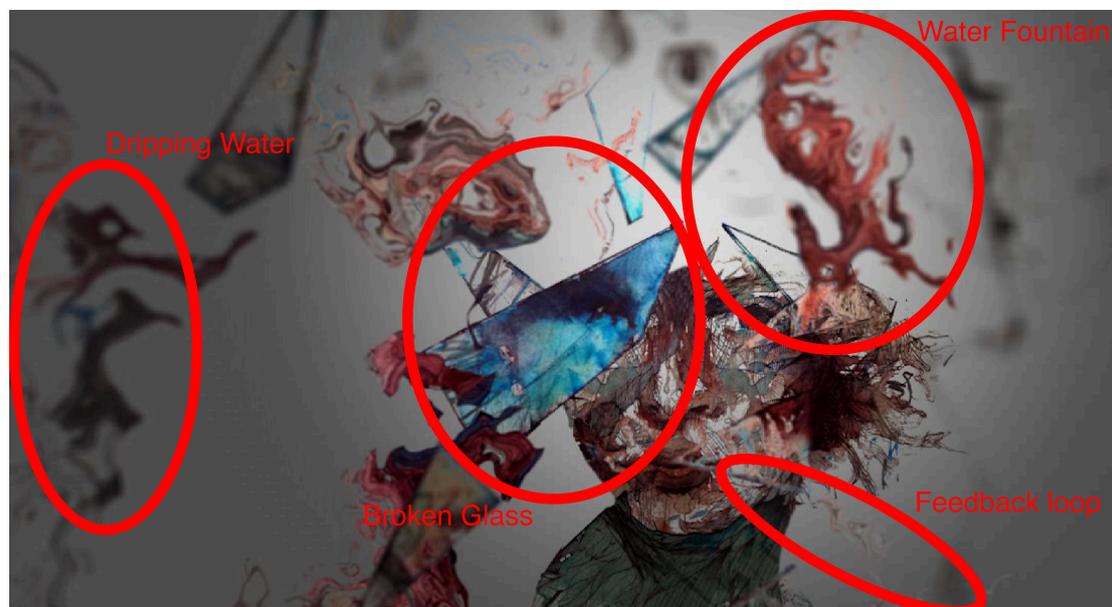


Figure 5.12. Sound Design of the piece *Another Place* (Griffiths, 2014).

This first experiment offers an indication of how sound can be used to shape, intensify, and enrich the observer's understanding of the virtual world.

Interactive Art Gallery exploits both sampled and synthetic sounds sources. In some instances, sampled sounds and recorded musical instruments are preferable, as they comprise audio recorded from real objects in real physical spaces. Therefore, they can be used to easily create complex sounds that would otherwise be extremely difficult to synthesise. On other occasions, synthetic sounds were considered to be more effective, as they can be entirely parameterised; which allows sounds to be synthesised to match motion parameters, providing important perceptible cues to the user regarding structural relationship between audio and visual objects.

In Interactive Art Gallery the participant explores the aesthetic experience by selecting the multiple visual elements presented on screen, that in turn control the series of audio loops. As the observer *hovers* the mouse over the elements displayed on screen, short percussive sounds are triggered to indicate that they consist of clickable objects (Figure 5.13).



Figure 5.13. Element of the piece *Vatican Slumber Party* (Lonsdale, 2014), is highlighted as mouse hovers over it.

The design principles and techniques applied in the development of Interactive Art Gallery consist of an important step in developing a system for connecting audio and visual objects.

5.7. Technical Evaluation

The techniques created and implemented during the development of Interactive Art gallery are significant to the development of the practical portion of this research in both conceptual and practical terms. That is, the techniques and approaches used in this collaborative project specify ways in which images can be analysed, explored and processed. As a complement, this chapter has also presented additional automated approaches that consist of algorithms that require minimal user interaction to perform each of the particular development steps. As a result, the design principles and techniques created during development of Interactive Art Gallery along with the additional automated approaches presented in this chapter, inform directions that were further explored and implemented in the creation of the algorithms presented in this research.

As mentioned earlier in this chapter, Interactive Art Gallery was implemented using Adobe Flash, because it appeared to be most appropriate solution at the time. The major concern voiced by the gallery at that time, was the potential problems regarding cross browser compatibility that could occur when the public tried to access the project through the website. Therefore, in order to avoid the risk of HTML code being interpreted differently by different Internet browsers, Adobe Flash was chosen as the platform for implementation.

However, multi-touch mobile devices are now one of the major channels of access to the web, and a great portion of these devices cannot access Flash content. Besides, the personal computers and mobile devices that are able to access Flash content are still required to have Flash Player installed in order to view and interact with the project. So, in the final analysis, the use of Flash can potentially become a barrier between the project and the audience.

Given these points, this research project has adopted the Processing integrated development environment as the platform for the development and the implementation of the software that will be presented as the practical

portion of this thesis. Processing can compile the code written in Java into JavaScript, and then, render the output in a HTML5 canvas element of a browser. Thus, it results in a web-based version of the project that can be accessed by any browser, including those found in mobile devices.

Interactive Art Gallery was presented during the 136th Audio Engineering Society Convention in Berlin, in April 2014, and received the Gold Award for Best Student Design Project (Refer to Appendix C). The outcome of this collaborative effort is fundamental for the development of this research project as it reveals creative steps that were later refined and applied to the development for AudioVisual Collage. For instance, the manually implemented image segmentation and inpainting process were evaluated and informed the creation of algorithms so as to automatically perform these tasks. My work as audiovisual designer in this collaborative project with artists from the Debut Contemporary art gallery has been highly productive. That is, this collaboration recognised important approaches that are part of my creative process and supported the creation of algorithms to automatically perform these techniques.

Therefore, they serve as an important guide for the development of the AudioVisual Collage project. Since the main purpose of the sound engine presented in this chapter is to create a correspondence between audiovisual objects textures and motions, it is also an intention of this project to investigate the level of details that a visual object requires from its auditory equivalent. This project attempted to create individual associations between visual elements and sonic elements, which were created to synchronously match. Nevertheless, these associations are sometimes difficult to be identified, due to the level of the narration, which is placed on the top of the backing track. Also, because there is not an opportunity for interaction where the viewer could directly operate the sonic texture.

Sound is hard to comprehend in the context of an application without observing how others react to it. The Interactive Art Gallery sonic design is a deeply enriching aspect of the audiovisual aesthetic experience. Therefore, it will undoubtedly demand a great amount of practice, testing, and tuning to the

accomplishment of its envisioned purposes in future developments. An important step that could be introduced to future developments of Interactive Art Gallery would be a user study, to observe how user react to the narratives and modes of interaction presented as part of the platform. That is, an important step that, due to the short time that was given to me to complete the project, could not be implemented. Thus, the platform could not be properly analysed from the perspective of a real user. Nevertheless, since it consists of an on-going effort, the user study will certainly be part of future developments. What is more, the Interactive Art gallery could be further explored by being applied to different art forms, including sculpture, dance and live performance.

Chapter 6 - AudioVisual Collage

“First we shape our tools, thereafter they shape us.”

(Marshall McLuhan, 1964)

6.1. Introduction

As previously discussed in the introduction and methodology chapters, AudioVisual Collage functions as both a study and a creative work. It consists of a screen-based interactive system that allows participants to create customised audiovisual aesthetic experiences in real-time, by combining digitised images and audio files of their choice. The system is comprised of a series of algorithms that automatically analyse and process images and audio so as to output an audiovisual aesthetic experience, exploring their audiovisual relationship. The algorithms include: image segmentation; inpainting; 3D composition; and a sound engine that extracts data from audio files and translates it into motion.

This chapter of the thesis outlines how the creative techniques applied during the collaborative creative process were recognised, refined and converted into a series algorithms that together form the AudioVisual Collage software. The software operates by analysing images and audio files selected by the participants. Next, the system uses the input to perform a data analysis that defines all other processes. For instance, the 3D lighting is controlled by the brightness mean value extracted from the images. That is, after segmenting the image the software calculates the image’s mean value in order to create the composition (foreground, middle ground and background). In order to create visual motion, the system performs FFT analysis to create a spectrum of the sound file input, and detects the frequency range that contains the highest amplitude. This approach is effective as long as there are not long periods of silence within the audio file. Later, data is extracted from this frequency range, turned into motion and applied to the separate layers created by the image segmentation algorithm (Figure 6.1).

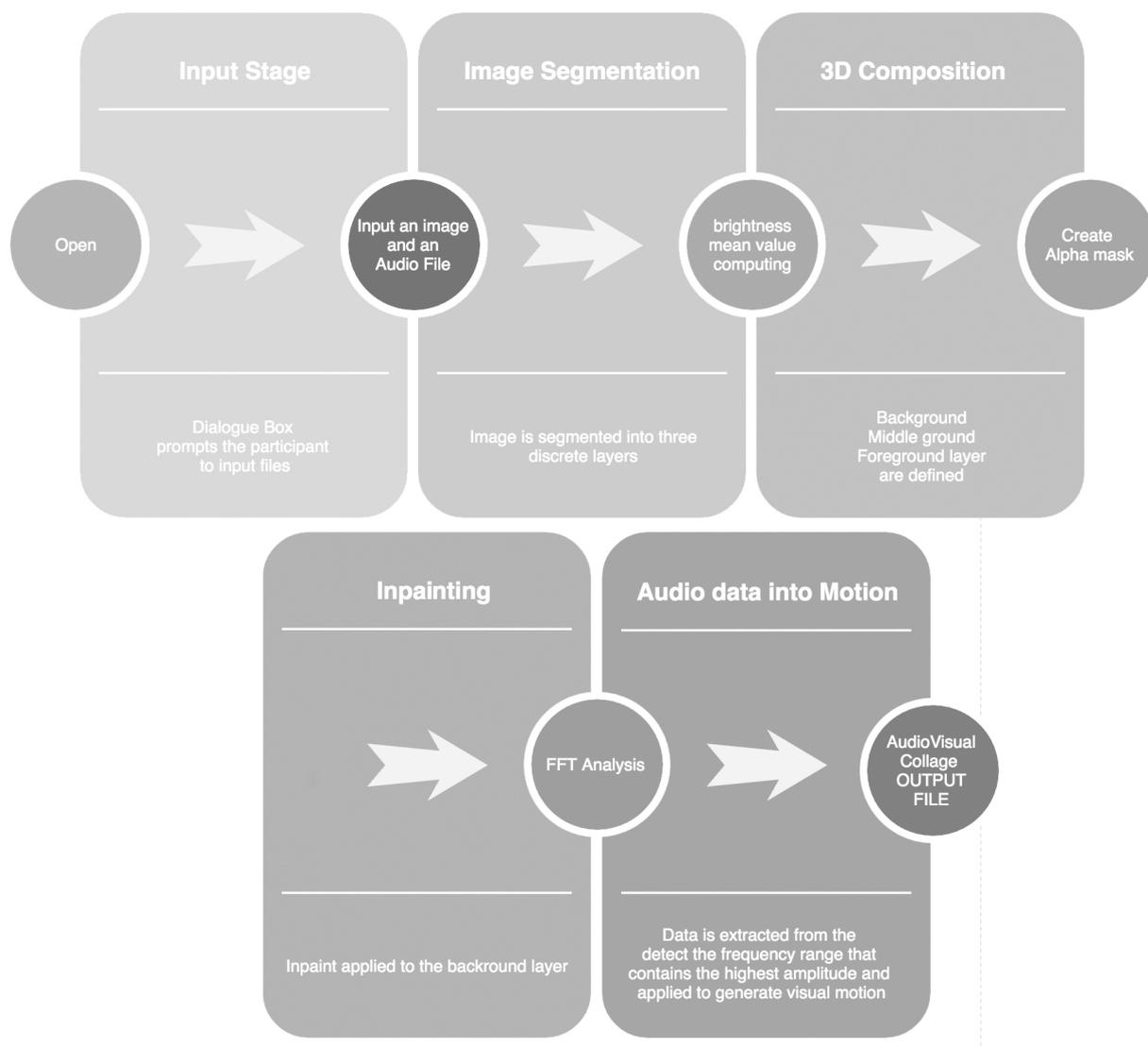


Figure 6.1. Illustration of the AVC processing.

AudioVisual Collage implements aspects and features informed by the creation of the two collaboration works displayed in chapters 4 and 5, as well as approaches and methods that have been outlined in the exploratory studies section in the methodology chapter. The practices and approaches described in this chapter consist of a design vocabulary, presented as the practical portion of my research. This design vocabulary was established by breaking down a complex system into manageable characteristics, for accurate and detailed analysis. The design vocabulary is embedded in the

AudioVisual Collage software and each one of its items comprises one of the software processing steps. These steps are: [1] the software prompts the participants to input an image and a sound file through a dialogue box; [2] an image segmentation algorithm splits the image into three discrete layers; [3] the system evaluates the image's brightness mean value to determine the layers composition (foreground, middle ground, background); [4] an algorithm applies inpainting to the background layer; [5] the system places the three layers within 3D space by applying the images as textures to quadrilaterals shapes; [6] the sound engine applies FFT analysis to the audio file to detect the frequency range that contains the highest amplitude and extracts data from that specific frequency over time; [7] an algorithm translates data extracted from the audio file and applies it to the transformation of position parameters of the visual layers to generate motion; [8] the system applies 3D lighting, where the light intensity is controlled by the results of steps 3 and 4; [9] the system implements a virtual camera that reveals the 3D composition and allows the participant to navigate with the use of the cursor; [10] the software plays the audiovisual aesthetic experience.

This chapter portrays all the aspects regarding the creative process of Audiovisual Collage. Initially, it presents the creation of the *image segmentation* and *inpainting* algorithm. Next, it describes the process of implementing these two algorithms in a prototype used for the user study, which is subsequently presented. Following this, this chapter defines how the feedback and insights, obtained in the user study, were evaluated and fed back to the development of a new prototype.

6.2. Project Concept

Once the key approaches that constitute my art practice were recognised, the next step was to determine how each of these steps could be automated and implemented as algorithms into a system designed to help other programmers, artists and designers to achieve similar results.

As outlined in chapter one, these steps are:

- **A method that combines automatic image segmentation and inpainting.**
- **A method for generating visual motion from audio data.**

The development of AudioVisual Collage software faced two main challenges: first, establishing a construction design able to materialise both practical properties recognised during the collaborative process as well as the theoretical aspects explored during this study; and second, identifying the technical requirements and skills necessary for the completion of the envisioned platform.

Multiple programming languages were evaluated in order to measure their suitability for each of the required tasks. The list programming languages tested include: Pure Data, Max/MSP/Jitter, Adobe Action Script Flash, openFrameworks and Processing.

After consideration, *Processing* (2016) was the chosen platform to implement the Audiovisual Collage prototype. The criteria utilised for the selection of the programming environment were threefold. Firstly, because Processing allows an easy adaptation of pre-existing source code resources into new algorithms, which could not be so easily achieved with some of the alternative programming languages considered for the project. Secondly, because Processing consists of an open source package, which makes it customisable. Thirdly, because apart from running in its own window display or compiled as OSX and Windows applications. Besides, Processing integrates itself seamlessly with Java, which allows code written within it to be easily incorporated into other software in future developments.

6.3 Developing The First Prototype

Two fundamental techniques applied in development of the collaborative projects consisted of *image segmentation* and *inpainting* processes. Therefore, in order to evaluate the implementation of these two techniques as algorithms, a prototype that automatically performs both tasks was built. The model, which is called *AVC Image Segmentation and Inpainting*, incorporates multiple automated steps: 1) it applies image segmentation to an image; 2) it splits the image into background layer, middle ground layer and foreground layer; 4) it applies inpainting to the background layer; 5) it saves the three layers as three separated files; 5) it allows the viewer to observe and navigate through the content generated within 3D space (Figure 6.2).

As previously stated in the methodology chapter, the idea of creating an early prototype was employed in order to create a model for evaluation. Thus, the results from this evaluation could be re-introduced into the development project as a way of further refining the project in a later stage. The study was conducted to obtain the views and opinions of other visual artists and designers after interacting with the system. Likewise, the study was created to discover if other artists would need or want such tool. Since the aim of this thesis is to deliver a design vocabulary in the form of a series of algorithms for art practitioners working in the field of interactive art, it was vital to obtain their evaluation with regards to the system's structure and features.

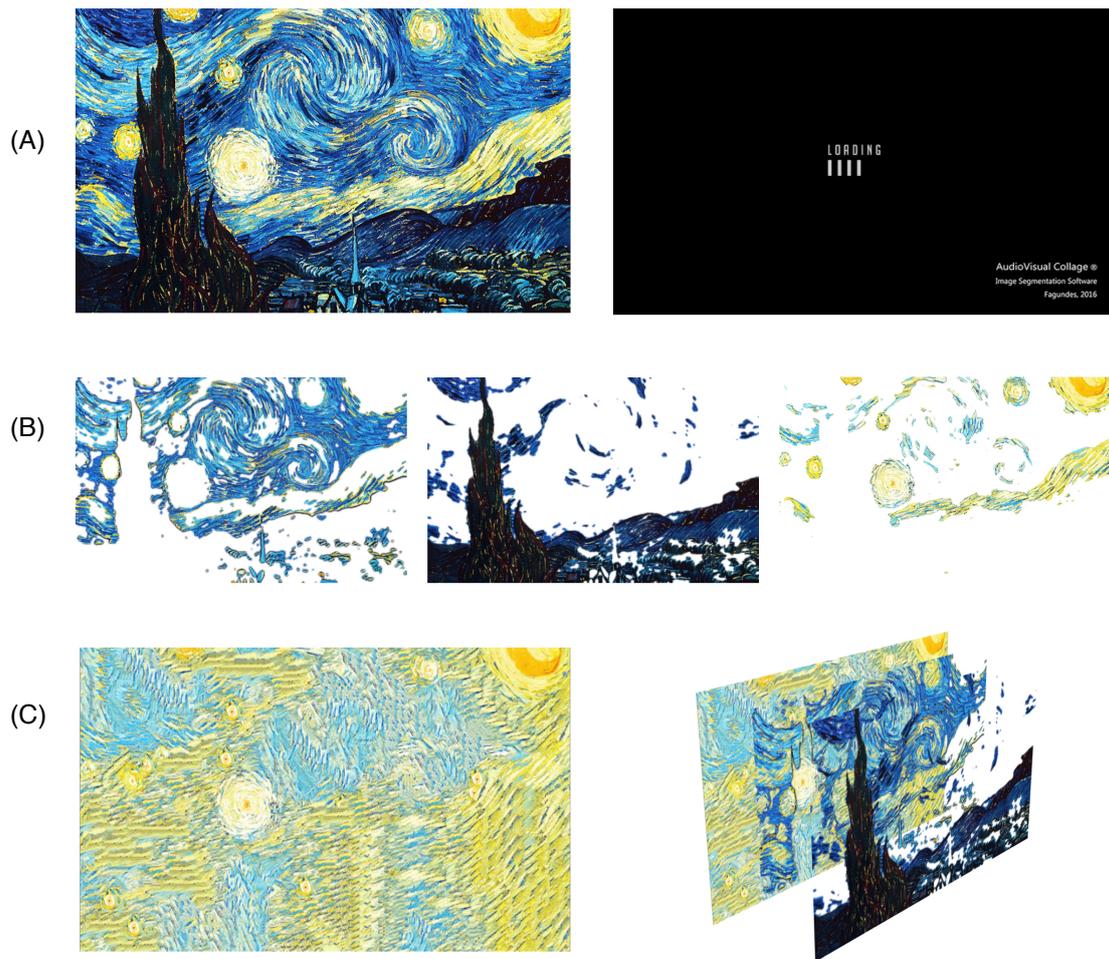


Figure 6.2. AudioVisual Collage visual processing (Fagundes, 2016).

(A) Image input by participant and analysed

(B) Image segmentation process

(C) Background Inpainting and 3D Composition process

The sections 6.4, 6.5 and 6.6 of this chapter outline the development of the algorithms that comprise the first prototype, which is followed by the user study described in details in chapter 7. The chapter 8 is devoted to the development of the final version of the AudioVisual Collage software, including its visual features as well as its sound engine. The section 8.9 presents a series of outputs generated with the final system in order to demonstrate its performance in a practical context.

6.4. Image segmentation algorithm

In computer vision, image segmentation comprises the process of splitting a digital image into multiple regions. Each single pixel found within those regions presents a certain sort of similarity with respect to some characteristic or computed property, such as brightness, colour, or texture. The result of image segmentation process is a collection of segments that collectively form a complete image (Shapiro and Stockman, 2001: 279).

In other words, the process of segmentation involves separating a given image into multiple regions. Segmentation of an image is usually performed through the identification of common proprieties within regions, or by identifying boundaries between regions, such as the edges of a shape. The simplest property that can be shared by pixels is intensity. Therefore, *thresholding* comprises a natural approach to segment such regions, as it creates binary images by turning all pixels below a specific threshold value to 0 and all pixels above the threshold to 1 (Morse, 2000).

In AudioVisual Collage the images are segmented into three different layers and placed within 3D space. The identification of this method consists of a valuable contribution presented by this thesis, as previously described in chapter one, because it consists of an approach that can be automated to allow other artists and designers to quickly achieve similar results.

AudioVisual Collage uses two different techniques to perform image segmentation. To explain, in some cases the images are split into two discrete layers, in those cases the *thresholding* technique is applied, because it creates binary images. In other instances, the images are split into three discrete layers, for those cases the *posterization* method is applied. Therefore, multiple techniques are applied to achieve specific results. In this section the two techniques are outlined and their implementation is explained in detail. It is important to clarify that these two techniques are not new inventions. That is, those algorithms consist of automatic segmentation libraries that were re-arranged and combined with a posterization method in a

new way. This proposed new configuration consists of one of the contributions of this research.

Furthermore, the techniques for improving the quality of image segmentation process are also outlined. It is important to note that there is no universal rule that can be applicable for image segmentation that could potentially work for any image; no segmentation technique is perfect. Thus, this project embraces multiple techniques in order to find different solutions for distinctive needs.

6.4.1. Thresholding Algorithm

Among the numerous image segmentation methods, *thresholding* (Shapiro, et al. 2001: 83) is arguably one of the simplest ones and perhaps, from a broad perspective, one of the most effective ones. From a grayscale image, *thresholding* creates a binary image. The Thresholding method replaces all pixel of an image, a white pixel if the image intensity value is greater than a predetermined value or with a black one if black pixel if the image intensity is less than a predetermined value (Shapiro, et al. 2001: 83). As the image below illustrates (Figure 6.3), the darker portion of the image becomes completely black, while lighter portion, which is comprised by the sky, becomes completely white.



Figure 6.3. (A) Original Image
(B) Thresholding algorithm applied

When the *thresholding* technique was applied to the prototype, it generated a crucial question: How to chose a threshold value that could be used for any image?

In other words, the *AVC Image Segmentation and Inpainting* application requests the participants to input a digitised image, which can potentially be of any sort, it might be representational or completely abstract imagery. Therefore, setting a fixed threshold value might not be a suitable choice because, in general, images have distinct brightness strength. To tackle the problem, and avoid having a fixed threshold value that might not fit different kind of images (at least partially), I used an adaptive threshold algorithm. That is, by adaptive means that the algorithm does not have a fixed threshold, but that the algorithm allows the threshold to vary depending on the image's brightness mean value (Figure 6.4).

Code:

```
SimpleGray image = Boof.gray(left, ImageDataType.F32);  
  
double mean_layer01 = image.mean();  
  
float meanValue_01 = (float)mean_layer01;
```

Figure 6.4. The algorithm that calculates the image's brightness mean value.

As the code illustrates the thresholding method iterates through all possible values and separates the pixels that either fall in foreground or background groups. The aim of using an adaptive threshold is to find a threshold value in which the sum of foreground and background is at its minimum (Figure 6.5). For instance, if an image contains large amounts of bright pixels, the threshold will be higher than one detected on a dark image.

Code:

```
// image's meanvalue is responsible for defining the threshold  
int threshold = meanValue;  
  
for (int i = 0; i < numberPixels; i++) {  
    pixelBright = brightness(image.pixels[i]);  
  
    if (pixelBright > threshold) {  
        // make it white  
        pixels[i] = white;  
    }  
    else {  
        // make it black  
        pixels[i] = black;  
    }  
}
```

Figure 6.5. Adaptive threshold algorithm.

6.4.2. Posterization Algorithm

The image posterization method consists of converting a continuous gradation of tones into regions of fewer tones, presenting abrupt changes between the tones. To explain, posterization converts an image into a new version that presents distinct, but flat areas of colours. A posterized image usually displays the same general appearance when compared to the original. However, parts of the images that previously contained gradual transitions become replaced by sudden changes in gradation from one tonality to another. In sum, the process consists of replacing entire groups of discrete gradations of hues into small number of very contrasting hues (Kwon and Chien, 2011). Image posterization was the method used to perform image segmentation on the images input by participants, while the thresholding technique was used for the creation of masks that function as maps to indicate the system where to apply segmentation.

Image posterization consists of an important topic in colour image processing, which has been vastly explored due to its effectiveness in porting images across devices. According to research by Kwon and Chien (2011), posterization can be applied to images that carry large number of colours in order to simplify them. Therefore, those images can be transported to devices that have restriction in the number of discrete colours hues that they are able to display. Likewise, it has been greatly explored as an artistic effect, which can be currently found in the most of the image editing software commercially available.

The posterization process reduces the number of colours of an image by individually reducing the number of distinct levels in each colour channel. Thus, the resulting image has less discrete values of Red, Green, and Blue. The posterization method takes four parameters: an RGB image object, its width, its height, and the posterization value. The later means the number of colours that the image will be reduced to. When the RGB values are specified, the image posterization algorithm calculates new R, G, and B values. As a result, the number of colours present in the resulting image is much smaller

when compared with the original. Consequently, it generates a poster image with the combination of these resulting colours (Kwon and Chien, 2011).

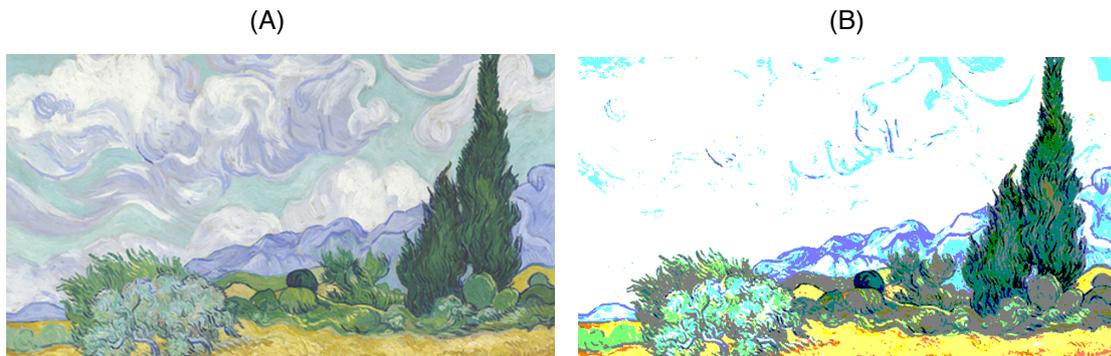


Figure 6.6. (A) Original Image
(B) Posterization algorithm applied

The posterization algorithm execution time is quite fast as it consists of a simple operation, and the resulting image presents a strong colour contrast, as the representative values includes both values 0 and 255. However, the algorithm produces some colours that would be difficult to identify (Figure 6.6). This initially seemed to be a rather a serious drawback of the algorithm. To explain, since the AudioVisual Collage image segmentation algorithm aims to split images into three fixed layers, it would be crucial to have the images reduced to three different, but constant, tonalities. Therefore, the solution found for this issue was to convert the image to grey scale before the process of posterization (Figure 6.7).

Code:

```
image.filter(GRAY);  
loadPixels();  
for (int i=0; i<image.pixels.length; i++) {  
int bright = int(brightness(image.pixels[i])/ meanValue) * 68;  
image.pixels[i] = color(bright);  
}  
updatePixels();
```

Figure 6.7. Posterization algorithm.

This approach resulted to be a successful solution for image segmentation because gray scale images contain only one channel. That is, one value for each of its pixels, which ranges from 0 to 255 of intensity information. Thus, the algorithm assures that any image, after converted into

grey scale, will be converted into a poster version containing only three discrete tones of grey (Figure 6.8).

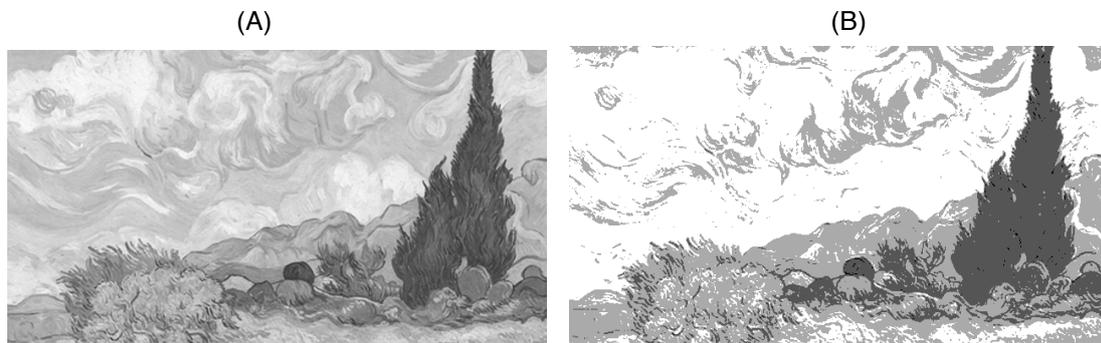


Figure 6.8. (A) Grayscale Image
(B) Posterization algorithm applied

Another issue found after implementing the posterization method, was that as the number of colours get reduced, some of the objects' shapes and edges are not well preserved in the resulting image, and some hard edges were produced, which generates a rough appearance.



Figure 6.9. (A) Posterization without Blur
(B) Posterization with Blur applied

The solution for this issue was to apply blur to the resulting posterized three coloured image, which generated a softened image with less noise and better defined edges in all three tone ranges before extraction (Figure 6.9). The blur achieves this by decreasing the quantity of high-frequency components in an image, working as a visual low-pass filter (Figure 6.10).

Code:

```
for (int y = 1; y < original_image.height-1; y++) {
  for (int x = 1; x < original_image.width-1; x++) {
    float sum = 0;
    for (int ky = -1; ky <= 1; ky++) {
      for (int kx = -1; kx <= 1; kx++) {
        int pos = (y + ky) * original_image.width + (x + kx);
        float val = red(original_image.pixels[pos]);
        sum += kernel[ky+1][kx+1] * val;
      }
    }
    newImg.pixels[y* original_image.width + x] = color(sum);
  }
}
newImg.updatePixels();
image(newImg, 0, 0);
}
```

Figure 6.10. Blur algorithm.

6.5. Inpainting Algorithm

Inpainting has recently become an active topic for research within the image processing field due to its challenging nature. The inpainting approach consists of algorithmically re-constructing corrupted or missing portions of a digital image. The inpainting method implemented in the AudioVisual Collage system was built using a library called PatchMatch (Barnes et al., 2009).

Barnes and his colleagues (2009) outlined, in their paper *PatchMatch: A Randomized Correspondence Algorithm for Structural Image Editing*, that there are three types of inpainting methods: image completion, image recomposing (also known as image reshuffling) and image retargeting. This research project concentrates on the image completion method, which consists of an algorithm that performs randomised searches for the nearest neighbour patches for completing the missing portions of an image.

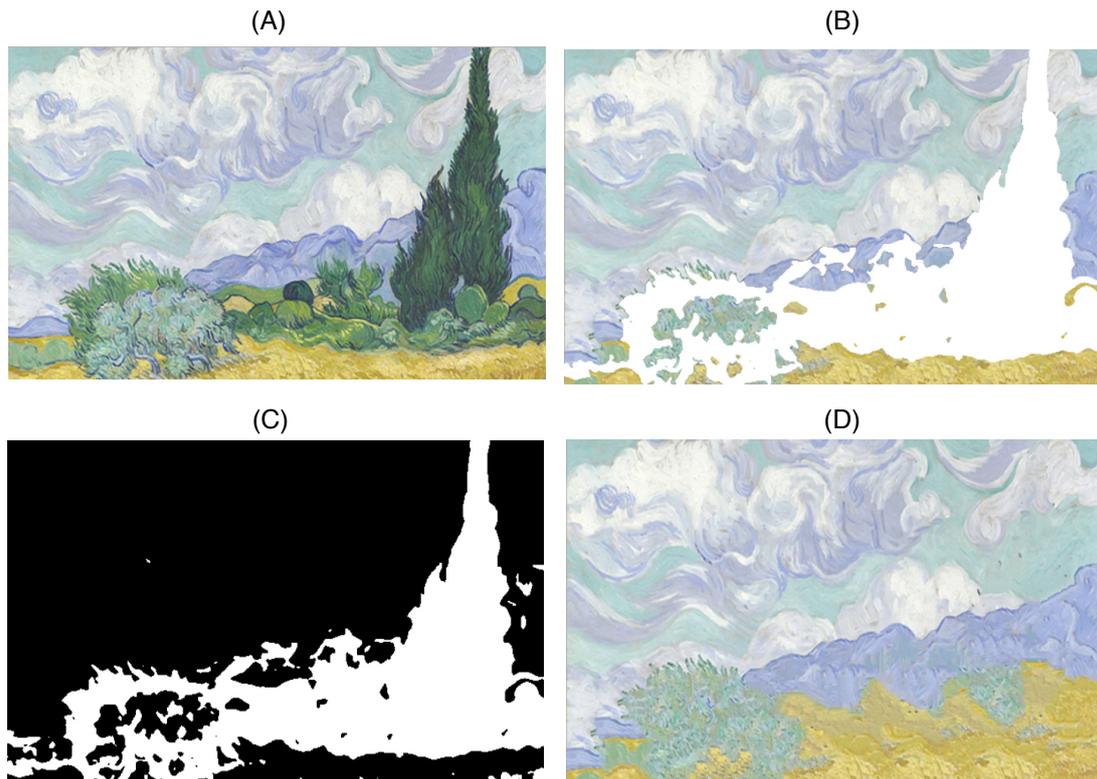


Figure 6.11. (A) Original Image
 (B) Image segmentation process
 (C) Image segmentation mask
 (D) Background Image with extracted area inpainted

The implementation of the inpainting method presented a crucial challenge. Most of the inpainting approaches, including the examples presented by the PatchMatch library (Barnes et al., 2009), require a considerable amount of user interaction. That is, they allow the user to draw curves across the missing regions or, in other instances, indicate from where to take and where to put texture. However, unlike most of these methods, the inpainting implemented in the AudioVisual Collage software automatically reconstructs the gaps left in the images by the image segmentation algorithm (Figure 6.11).

The implementation of an automated approach using the already existent library was performed by analysing the gap left in the image after the image segmentation method and converting this region into an alpha mask,

which works as a map to inform the algorithm the coordinates of the area to be inpainted (Figure 6.12).

```
Code:  
  
// the area segmented is turned into a binary image  
image_segmented_mask(THRESHOLD,0.1);  
  
patchMatch=new PatchMatch(this);  
  
//the segmented image is turned into a map and applied  
to define the area to be inpainted  
  
patchMatch.patch(inpaint_original, image_segmented_mask,1);
```

Figure 6.12. Turning an image section into a binary image.

As research by Bernes et al. (2009) points out, image completion of large missing regions is always a challenging task, as even the most advanced methods can still produce inconsistent results. Therefore, in the case of an image that has nearly half or even more of its content extracted, the algorithm can sometimes generate surprising artefacts. Nevertheless, it is important to keep in mind that the inpainting process is performed in the background layer, which is a portion that is, during most of the time throughout the audiovisual aesthetic experience, partially obstructed by the middleground and foreground layers.

A fundamental question emerged after the image segmentation process: which of the three layers comprises the image's background? As the question suggests, some sort of analysis on the images was necessary in order to distinguish between the objects of interest contained in the image and the background. The technique used to determine the 3D composition, that is, which layers consists of background, middleground and foreground, was to obtain the mean value of the image. Providentially, this task has already been performed by the system during the thresholding process, as described in section 6.4.1. Therefore, the next step is to set a condition that would work as a global average for any image input into the system. To explain, pixels

values consist of integers that range from 0 to 255, the later being white and 0 being black. As already described in section 6.4, all images were converted to grey scale during the image segmentation process. For an 8-bit gray scale image there are 256 possible intensities. Therefore, the mean value algorithm goes through all pixels contained in the image and calculates the intensity of each one, in order to provide the average intensity of all pixels. As a result, any image with brightness mean value less that 127 would have its darker layer as the background, and all the images that have its brightness mean value equal to 127 or higher, would have the lighter portion as the background.

The image below (Figure 6.13) displays an image that presents a brightness mean value of 138.4. Thus, since the brightness mean value is higher than 127, the algorithms classifies the brighter portion of the image as background and applies inpainting to it, over the area extracted during the image segmentation process.

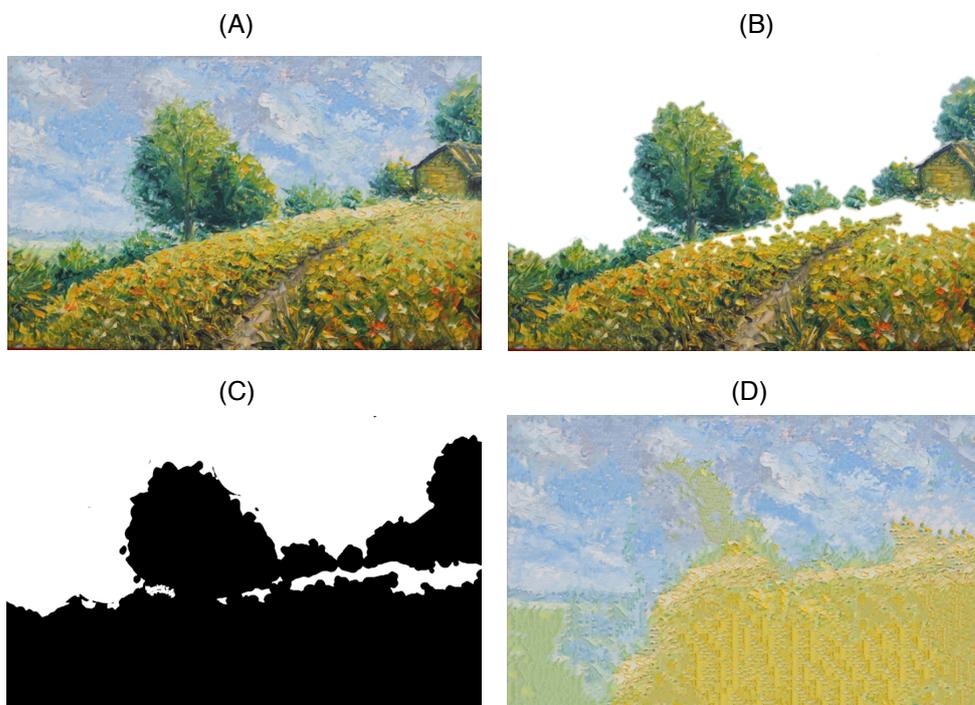


Figure 6.13. (A) Original Image with brightness mean value of 138.4
(B) Area of the image segmented
(C) Image segmented
(D) Background Image inpainted

The example below presents an image (Figure 6.14) with brightness mean value of 26.31, in that case, the inpainting algorithm classifies the darker portion of the image as background and applies inpainting to it.

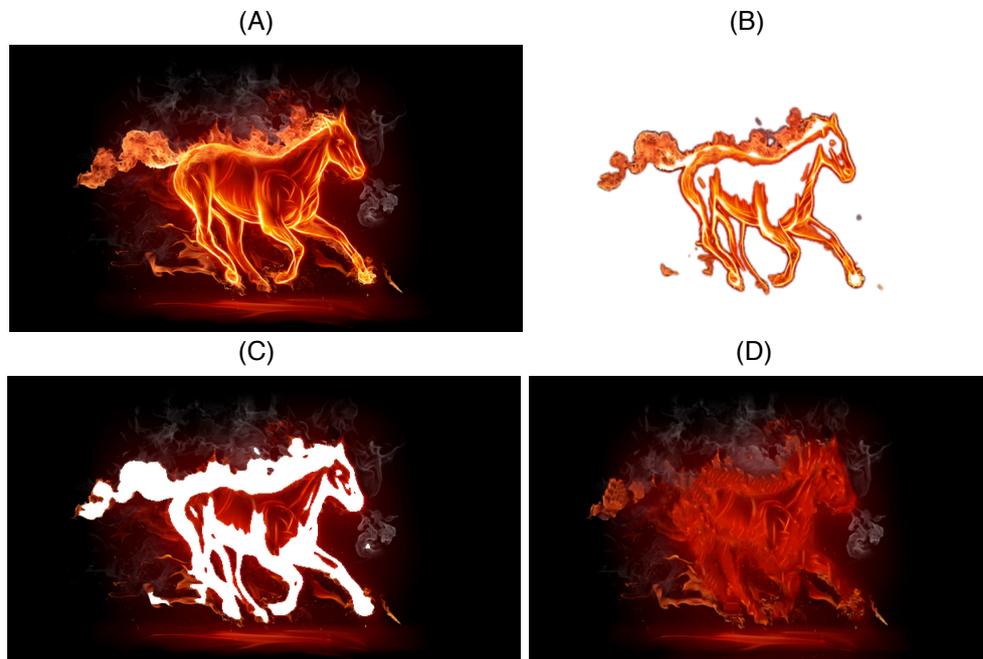


Figure 6.14. (A) Original Image with brightness mean value of 26.31
(B) Area of the image segmented
(C) Image segmented
(D) Background Image inpainted

As shown above, the adaptation of the PatchMatch inpainting algorithm presented in this section is arguably a novel approach for computing approximate nearest neighbour fields for sections automatically extracted by an image segmentation algorithm. This inpainting method delivers an outcome fast enough that it can be executed along with a series of other processes simultaneously.

6.6. 3D Composition

Once the segmentation process is completed with the implementation of both *thresholding* and *posterization* processes, as described in section 6.4 of this chapter, the separate layers are individually placed within 3D space.

The technique used to determine the 3D composition, that is, each of the layers consists of background, middle ground and foreground, was to obtain the mean value of the image as described in section 6.4 of this chapter. In order to place the three separated image layers within 3D space, it is necessary to create three separate quadrilateral shapes. In Euclidean plane geometry, quadrilateral (also known as quadrangle or tetragon) consists of a polygon with four sides and four corners (Johnson 1929, p. 61).

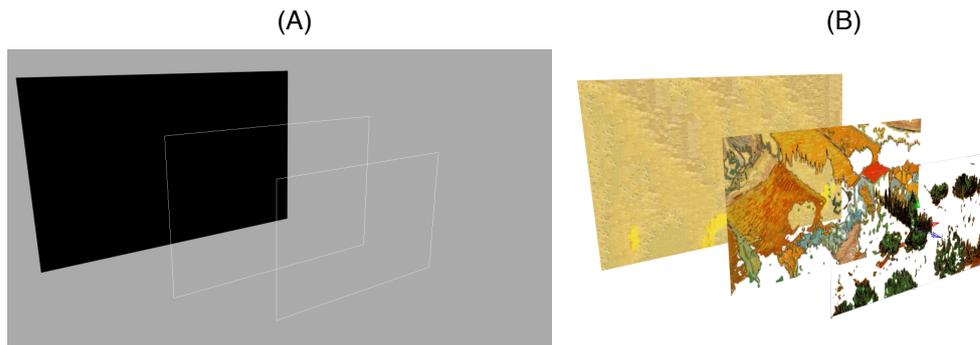


Figure 6.15. (A) Three quadrilateral shapes
(B) Images applied as textures to the 3D shapes

The image above (Figure 6.15) illustrates the images applied as textures to the quadrilaterals with four vertices, and with a transparent fill. The `texture()` function states the texture image to be used, while the `vertex()` function maps the image to the geometric shape. It is important to notice that the images were segmented into multiple layers, as outlined in section 6.4. Therefore, the layers were individually applied to quadrilaterals so as to freely and individually move them within 3D space.

```
Code:
beginShape();
tint(255,225);
texture(image);
vertex(0, 0, 0, 0, 0);
vertex(709, 0, 0, image.width, 0 );
vertex(709, 443, 0, image.width, image.height);
vertex(0, 443, 0, 0, image.height);
endShape();
```

Figure 6.16. Images applied as texture to quadrilaterals generated within 3D space.

The above code consists of a very simple implementation of texture, as it purely consists of an image applied to a shape created from its own dimensions (Figure 6.16). However, this simple approach generated numerous possibilities for the final implementation of AudioVisual collage.

The AVC first prototype allows the participant to observe the outcome of the image segmentation and inpainting processes in a three-dimensional environment, through the implementation of a virtual camera. The camera is directly operated in real-time through the user interface, based on mouse X-axis and Y-axis position, to ensure the means of controlling the outcome is as intuitive as possible. The system allows small changes in the camera position in order to apply parallax movements to the images positioned in virtual space and to provide the participant an opportunity to dynamically explore the results (Figure 6.17).

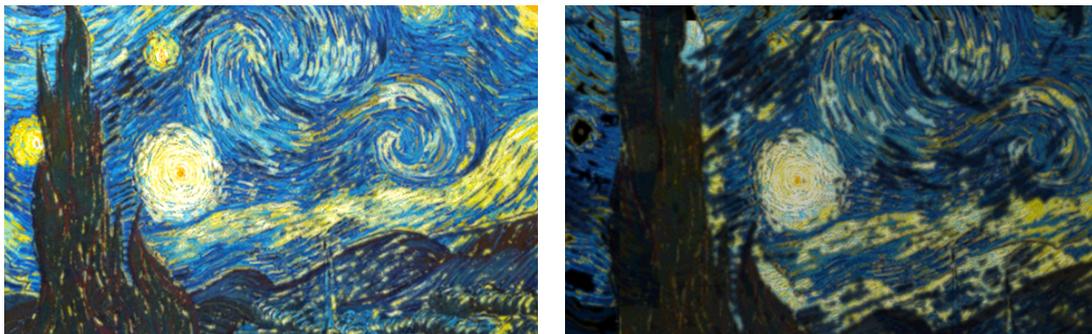


Figure 6.17. Visual Layers placed within 3D space to allow the participants' exploration.

The *AVC Image Segmentation and Inpainting* prototype has only one camera, which was implemented for the purposes of visualisation. The idea was to create an interface to stimulate participants' curiosity and encourage them to explore the layers within 3D space. This follows the approach suggested by Donald Norman (2002: 183) with regards to explorability, in which he proposes that an effective method for making a system easier to learn, and to use, is to make them explorable. According to Norman, it can be achieved by motivating users to learn and experiment the possibilities through actively exploring the system.

At this point, the *AVC Image Segmentation and Inpainting prototype* had achieved a substantial level of development. Therefore, in order to assess whether the project development was going in the right direction, a user study was created so as to generate feedback from other artists and audiovisual designers regarding my technical choices and artistic decisions.

Chapter 7 - User Study

“Design and art are independent coordinates that provide their greatest satisfactions when experienced simultaneously”

(Milton Glaser, 1943).

The study was conducted with ten participants with background in digital arts, including: Computational Arts, Creative Computing, Digital Cinema, Audio & DSP Engineering, and Web Design. Since the system was going to be evaluated by artists and audiovisual practitioners with substantial knowledge and experience working with interactive systems and applications, it was necessary for the system to have achieved a sufficiently advanced state of development to cope with their demands. Therefore, the system went through a series of design iterations to guarantee it would work correctly during the study. The complete designing of the prototype was outlined from chapter 6.1 to 6.6.

The *AVC Image Segmentation and Inpainting* prototype is arguably a novel approach, proposing image segmentation and inpainting processes in a unique and completely automated system. The focus of this study is to present users with a prototype that automatically performs the tasks of image segmentation and inpainting. The prototype was presented to professional art practitioners, so as to gather information to achieve the right design. The survey was expected to recognise positive and negative aspects of the prototype, as well as opportunities and areas of interest to be further developed and refined.

7.1. Method

The user study was based on a questionnaire comprising eighteen questions, which were responded to by ten participants. The participants are named P1 to P10. Six of the ten respondents (P1, P2, P4, P6, P7 and P9) were male, and four respondents (P3, P5, P8 and P10) were female. The decision to have only a small number of expert participants and to create a

relatively short questionnaire was taken to allow the collection of qualitative data and to focus on the most important aspects of the project in an early development stage.

The participants received a folder containing the *AVC Image Segmentation and Inpainting* application, the questionnaire and a text file that briefly explained the project. The participants were given two tasks. Firstly, to load an image of their choice and to observe the outcome of the image segmentation and inpainting processes. Secondly, to answer the questionnaire provided, which is presented in Appendix D. The participants were encouraged to access and operate the system at any time during a period of two weeks. The idea was to provide the participants with sufficient time to repeatedly experience the system and to reflect on their experience in order to accurately respond the questionnaire. The application was build in Processing and compiled as a Java application compatible with Macintosh and Windows platforms.

The questionnaire focuses on the fundamental research topics investigated in this thesis: aesthetic experience, content and interactivity. The survey was intended to gain insights from visual artists and professional designers with regards to the artistic and technical decisions that I have made during the iterations of the *AVC Image Segmentation and Inpainting* creative process, as informed by the case studies (specifically case study 2). This study helped me to form a new understanding of the system by revealing its positive and negative aspects, as well as collecting suggestions on how to improve it in the final version of AudioVisual Collage. Once all the responses from the user study were collected [Appendix D], participants were asked to undertake a second user study. The User Study 2 is described in section 4 of this chapter.

7.3. User Study Results

All the artists and designers invited to take part in the study are professional art practitioners, which work in the field of interactive art. Three participants (P2, P5 and P6) have experience as audiovisual designers; three

participants (P4, P8 and P10), as visual artists; and the others (P1, P3, P7 and P9) have experience as both, artists and designers.

With exception of one (P8), all participants have previously used image segmentation and inpainting (or texture synthesis) tools before. Seven participants (P1, P2, P4, P5, P6, P7 and P9) indicated Adobe Photoshop as the tool previously used, while two participants (P3 and P10) pointed out Adobe Lightroom. Both tools mentioned by the participants are tools that require user interaction to perform the image segmentation and inpainting tasks, thus, they were all using an automatic system for the first time.

With regards to foreground/background separation, the majority of users (7) are extremely satisfied with the results, while the other participants (3) consider that they are simply satisfied. Regarding the composition generated by *AVC Image Segmentation and Inpainting* prototype, that is, the elements that the application recognises as background, middle ground and foreground, six participants (P1, P2, P5, P6, P8 and P10) were extremely satisfied with the content generated while three (P3, P4 and P7) were very satisfied and one (P9) was satisfied. Since the nature of the images input by the participants was completely unpredictable, the results obtained in the study indicate that the algorithm is fairly consistent in terms of foreground/background separation.

According to the study, the inpainting task was successfully accomplished, as the majority of the participants (6) pointed out that they are extremely satisfied, while the two participants (2) were very satisfied and one was satisfied.

With regards to manipulation options, most of the participants (6) considered an automated application (with no adjustable parameters) suitable for the task of segmenting still images to be very appropriate. While, two participants (P2 and P6) considered it to be very appropriate, one participant (P10) considered it to be appropriate. Likewise, most of the participants (6) considered that they were extremely satisfied with an application that combines image segmentation and image inpainting in an automated system, while three (P6, P8 and P10) were only satisfied.

When asked if they would like to have control over the number of layers resulting from the image segmentation process, the participants' responses were divided. Half of the participants (P1, P4, P5, P8 and P10) responded that it would be extremely important to have control over the number of layers generated by the image segmentation process, one participant (P6) considered that it would be very important, while two participants (P3 and P7) considered it would be important, and two participants (P2 and P9) considered it not important. Similarly, regarding the user interface clearness, only two of the participants (P3 and P6) consider that the user interface is extremely clear. Besides, one participant (P2) has pointed out that:

"...the addition of a visual indicator to clarify how the position of the mouse affects the image could be useful."

Eight participants (P1, P3, P4, P5, P6, P9, P8 and P10) responded that they would use *AVC Image Segmentation and Inpainting* prototype as a tool in the creation of their own art practice. When asked to outline the reasons why, one of the participants (P4) outlined that:

"It's practical and effective and it helps the artist to access the image from other perspective".

Other participant (P8) responded that: *"Because it is easier to analyse separately the parts of the image you'd like to"*.

Other participant (P10) pointed out that:

"it is an interesting tool if developed further".

The user interface of *AVC Image Segmentation and Inpainting* application allows the participants to explore the content generated by both image segmentation and inpainting processes. Although the movement offered by the display user interface was only intended for the purpose of visualisation, one of the participants (P4) pointed it out as being an interesting feature of the system. In the participants' words: *"I like the surreal effect it does in the images, and also the possibility to give movement to the photos"*.

When participants were asked if they felt they were creating their own work while operating the prototype, six of the participants (P1, P3, P4, P6, P7 and P10) responded 'yes', while three participants (P2, P5 and P9) responded that they were not totally confident they were creating their own artwork, and one participant (P8) answered that she did not feel she was creating her own work. Considering the contrasted answers to this question, it is evident that more could be done in terms of customisation of the content, functionalities and amount of manipulation options offered by the system. One of the participants (P9) provided a comment that serves as an important clue to help clarifying this result:

“the app could offer a “zoom” tool allowing the user to add or subtract the amount of details to be extracted. In that way, it would feel more like doing my own artwork.

Since the *AVC Image Segmentation and Inpainting* application is being proposed as a tool to enable other artists to create their own artwork, the participants were asked if the images generated by the image segmentation and inpainting process inspired their creativity; the majority of the respondents (7) responded 'yes', while the others (3) responded that they did not. When asked why they felt inspired or not inspired, one participant (P4) stated: *“I think from the results of the applications the creativity starts”*. Other participant (P7) stated: *“Because it gives more possibilities for image editing.”*

The study also enquired if the participants considered that *AVC Image Segmentation and Inpainting* application would be suitable as a plugin for graphics editor software (eg. Adobe Photoshop); most of the participants (8) responded that it would be extremely suitable, while the others (2) considered to be very suitable. The study concluded with a final open question, which requested the participant's suggestions on how the prototype could be improved. That turned out to be a very interesting portion of the questionnaire. Since all the participants were skilled artists and designers, working professionally in the audiovisual and interactive art industry, they were able to provide valuable knowledge and remarkable insights. The suggestions were

evaluated and employed in the next iterations of the development process.

One participant (P1) stated:

“I think the application itself is easy to use and very straightforward, the suggestion I have does not involves the application itself but what could be developed around it, for example, a short demo video explaining what you can get from the application”.

Other participant (P3) made a comment regarding the inpainting process: *“Threshold and background reconstruction could be improved”*. One

participant (P9) left a very interesting comment regarding the user interface, he suggested:

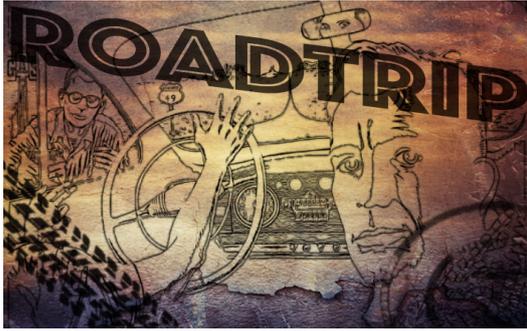
“...the addition of a visual indicator to clarify how the position of the mouse affects the image could be useful.”

7.4. User Study 2

Once all the responses from the user study were collected, participants were asked to undertake a second user study. The User Study 2 consisted of creating a piece of work utilising the images generated by *AVC Image Segmentation and Inpainting* prototype. The User Study 2 was undertaken by three participants (P2, P5 and P10). The participants were given completely freedom to choose the theme and the amount of images to be used in the creation of the visual pieces (Figure 7.1).

The idea behind the User Study 2 was to discover: 1) how other artists and designers would use *AVC Image Segmentation and Inpainting* application in their own art practice; 2) to what extent they would use it: and 3) in what context. The resulted artworks show how each of the participants has incorporated the system in their pieces. As it can be seen in the artworks above, the participants have employed the images generated by the system in different ways and in different degrees in the creation of their artworks.

(A)



(B)



(C)

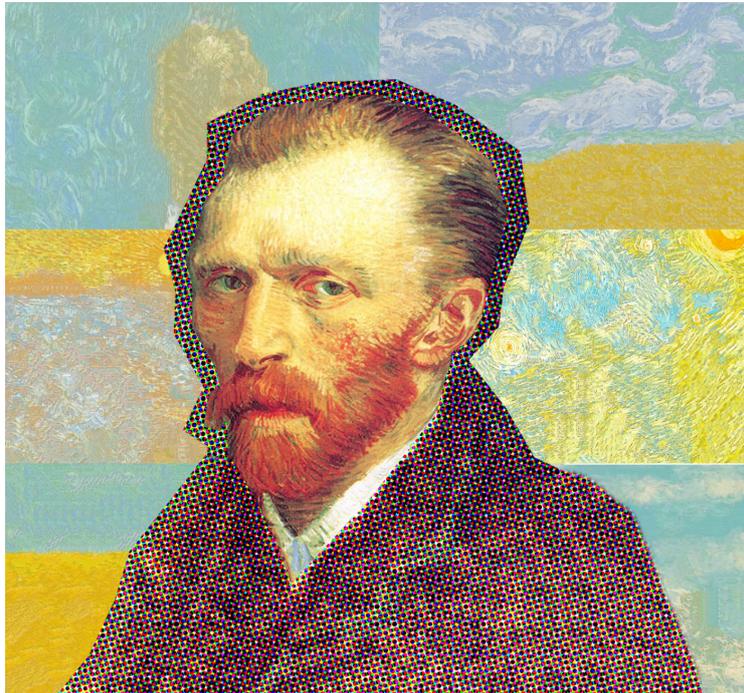


Figure 7.1. (A) *Artwork produced by Participant 2*
(B) *Artwork produced by Participant 8*
(C) *Artwork produced by Participant 1*

7.5. Discussion

Practice-based research endorsed me to apply my artistic approaches and techniques in the development of the collaborative projects, as described in chapters 3 and 4. Furthermore, the collaborative process allowed me to identify the essential techniques that constitute my art practice, which could

be automated with the creation of algorithms to assist other artists to create their own artworks.

During the development of the collaborative projects, I took in account my own perspective as both artist and observer. The user study permitted me to compare my own initial viewpoints with the opinions of other audiovisual designers and art practitioners. Some of the initial assumptions were confirmed and others were not, as a result, very useful insights were generated.

The outcome of the user study shows that the *AVC Image Segmentation and Inpainting* prototype has successfully achieved a number of its goals. For instance, the content generated by the system has consistently inspired the participants' creativity. Moreover, the vast majority of participants agreed that *AVC Image Segmentation and Inpainting* application is suitable to be implemented as a plugin for a graphic editor, such as Adobe Photoshop. Likewise, the study reveals that all participants would use *AVC Image Segmentation and Inpainting* application as a tool for the creation of their own artworks. The images used by the users can vary from figurative to abstract visuals, as well as from high to low definition. Therefore, based on the study's outcome, in terms of foreground/background separation and overall composition, the proposed algorithm presents consistent and accurate results when evaluated via qualitative methods.

Yet, the study also shows areas in which the proposed system could be improved and further developed. Participants have pointed out specific issues regarding usability in different areas of the questionnaire, including suggestions on how to improve the system. Some participants were very satisfied with a completely automated system that combines image segmentation and inpainting process, while other suggested that they would prefer to have more control over the generated images. Similarly, some respondents value a larger set of controls, while others participants show preference for a smaller amount of manipulation options. This underlines the relevance of user experience studies in the development of an interactive art project, as suggested by Höök et al. (2003).

The questionnaire included enquires concerning content, interactivity and experience; thus, the results of the study provided me with a wide range of insights that will be fundamentally useful for future development. The study has also served to investigate the project in new ways, which were not anticipated. One example of this could be observed in a suggestion made by one participant (P4): "...I like the surreal effect it does to the images, and also the possibility to give movement to the photos". This suggestion referred to the system's visual display, which was intended to only reveal the outcome of the image segmentation and inpainting processes within 3D space. Nevertheless, the suggestion made by this participant inspired me to emphasize and expand the movement generated by the image display in the AudioVisual Collage project.

The User Study 2 has shown how other artists employed *AVC Image Segmentation and Inpainting* application in the creation of their own artworks. One participant (P2) has created a colourful piece composed of abstract imagery. The other two participants (P5 and P8) have produced pieces that contain figurative images combined with textures, and typography. The variety of themes and approaches presented in the participants' artworks show that the *AVC Image Segmentation and Inpainting* tool successfully adapted to distinct scenarios. The additional task has also revealed the clear potential of the proposed system to empower other artists to create their own artworks.

When compared to the approaches used during the collaboration projects, where those techniques were performed manually, the algorithms that comprise the prototype prove to be efficient. Nevertheless, the rules and constraints that define the way in which the prototype operates, sometimes generate unwanted and unexpected results. That means that further refinement and user test is required. Nevertheless, the purpose of this study was to enquire if other designers and artists would want to use a tool to automatically perform the tasks of image segmentation and inpainting simultaneously. The result was positive; therefore, it encouraged the implementation of a new prototype that will result in the final version of the AudioVisual Collage.

Chapter 8. AudioVisual Collage Final Prototype

“An important technology first creates a problem and then solves it”

(Alan Kay, 1982)

As previously indicated, the first prototype was presented as a vehicle to evaluate the implementation of the image segmentation and inpainting techniques as algorithms. The outcome of the user study provided me with a new perspective regarding my own work as well as with valuable viewpoints and opinions that were implemented during the development of the AudioVisual Collage final prototype.

As the introductory section of this chapter outlines, AudioVisual Collage consists of software designed to allow artists and audiovisual designers to create their own audiovisual aesthetic experiences. This section summarises in which areas the first prototype successfully achieved its goal and what are the aspects that can be improved and refined in the final model. The system aims to offer a novel automated approach for the creation of audiovisual aesthetic experiences. That is, AudioVisual Collage extracts information from images and audio files and uses that information to control motion, lighting and a sound engine.

After being processed by the *image segmentation* algorithm, the images are applied to 3D quadrilateral shapes as textures. From that moment, they are converted into 3D objects and become part of the AudioVisual Collage’s 3D environment. The technique used to determine the 3D composition, that is, which layers consists of background, middle ground and foreground, was to obtain the mean value of the image. In addition, AudioVisual Collage exploits a technique that uses information from the image’s brightness so as to control the 3D lighting effect.

This technique helps integrating and blending together the three-dimensional space, so as to avoid a harsh juxtaposition of shapes and

colours. The implementation of this technique is outlined in detail in the next section of this chapter.

8.1. Using Brightness to Control 3D Lights

Once the images are segmented and the layers applied to the quadrangle 3D objects as textures, it is possible to observe them positioned within 3D space through the simulated lens of the virtual camera. At this stage, artificial lighting effect is used to illuminate these objects in the 3D environment. AudioVisual Collage segments 2D images, split them into separate layers and place them within 3D space. Then, it attempts to reconstruct the same lighting characteristics of the 2D image into the new 3D representation. Hence, the 3D lighting effect plays a crucial role in maintaining the overall scene consistency in terms of luminosity.

Additionally, the 3D lighting effect is responsible for increasing the general sense of depth and separation between the segmented layers, adding a dramatic fill to the 3D scene (Figure 8.1).

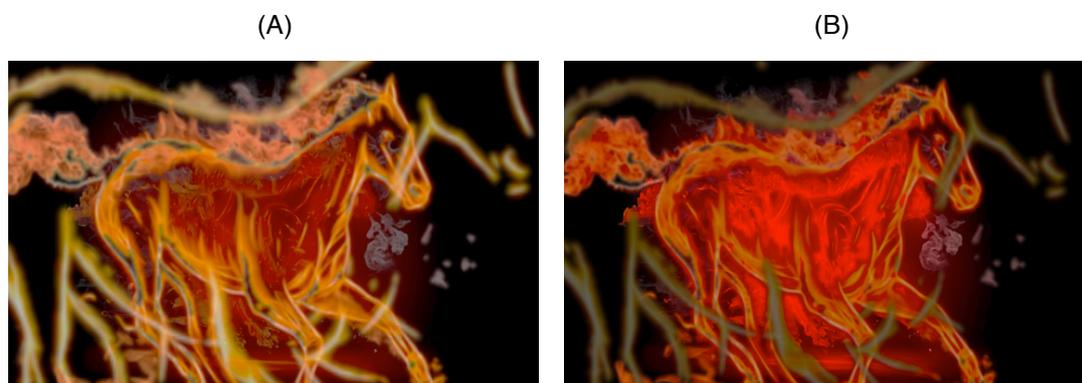


Figure 8.1. (A) 3D space without lighting effect
(B) 3D space with lighting effect applied

The system starts by using the image's brightness mean value, which was already calculated to determine the 3D composition during the image segmentation process and applies it to the direct control of the 3D lights. As described in chapter 6.6, the brightness mean value of the image defines whether the background will consist of the darker or the lighter portions of the

image. Thus, following the same notion, the lights intensity will vary according to the mean value of the image (Figure 8.2).

Code:

```
// 3D Lighting applied to the a composition with brighter background
if (meanValue1 >= 127){
    ambientLight(120,120,120);
    pointLight (255 , 255, 255, 709/2, 443/2, -100);
}

// 3D Lighting applied to the a composition with darker background
if (meanValue1 < 127){
    ambientLight(40,40,40);
    //pointLight(255 , 255, 255, 709/2, 443/2, 200);
```

Figure 8.2. Image's brightness mean value controls the light intensity.

To illustrate, if the image's mean value is higher than 127, the brighter portion of the image will comprise the background. Thus, the lighting directed into the background layer will have, in its initial position, a higher intensity in the background image (Figure 8.3).



Figure 8.3. (A) 3D space with lighting effect applied

AudioVisual Collage aims to find balance between intricacy and readability. That is, an image that plainly displays its shapes and that is

equally lit across might be easier to understand. However, it is not a very interesting image when compared to another that falls in and out of light and present a strong sense of depth. Similarly, images that present a clear definition between foreground/background might be easier to read. Yet, it does not present a particularly interesting spatial organisation. As researcher Calahan (1996) suggests, the human brain is remarkable at identifying forms and shapes with only a minimum amount of information. Therefore, a visual object lit with only a minimal amount of light can excite the viewer's imagination and evoke creativeness.

The 3D lighting effect resulted to be a fundamental feature of the AudioVisual Collage project. The implementation of lighting effect contributes to the notion of emulating processes present in the real world within the digital domain, and allows this research project to explore it in practical terms.

The implementation of 3D lighting was split into two parts: `ambientLight()` and `pointLight()`. The ambient lights can be placed anywhere within 3D space as they control the overall light and tone of the scene, while the point lights, which focus on specific layers, are placed in precise locations within 3D space so as to provide them with customised illumination.

The ambient light creates a basic continuity throughout the journey, affecting the overall tone and coloration of the world. The intensity of the ambient light is scripted in such a way that it modulates its intensity based on the virtual camera positioning. It varies from 0% to 100%, with full illumination applied when the camera is at its further point from the images. This is very important since the images that are used by the participants are unpredictable. Thus, the ambient light is an efficient tool to sustain continuity between the multiple cameras and to preserve the overall visual consistency, in terms of tone and colour throughout the scene.

The point lights provide individual illumination to each element within 3D space, dynamically changing its intensity based on the image's brightness mean value. Following the *perceptual grouping* principle, which is grounded on Gestalt psychology, as outlined in chapter 2.4, the 3D lighting effect was implemented in this project to help unifying a possible disorganized and busy

scene created by the image segmentation and inpainting algorithms. As previously stated, another reason for implementing 3D lighting effect was to increase the sense of depth and separation between the segmented layers of the images.



Figure 8.4. 3D lighting effect emphasizing the brighter background

As the image above illustrates (Figure 8.4), the scene presents an image segmented comprising three layers that are placed within 3D space, from back to front: the night sky, the town lights and the tree. Lighting each one in reference to each other enhances depth as each layer is clearly defined by the different brightness.

By successfully illuminating a scene it becomes easier to visually organise it. In order to successfully direct the viewer's eye, an image needs a focal point. Otherwise, the viewer's has no reward for having tried to look to a specific point on the image. Therefore, by establishing focal points using lights intensity and placement, it is possible to draw the viewers attention and give the viewer an opportunity to visually explore the space, without overwhelming them with excessive visual information. The techniques presented here are not strict rules, but consists of a suggested structure by which to explore artistic and creative possibilities.

8.2. The Sound Engine of AVC

This section presents the design decisions made during the development of AudioVisual Collage's sound engine as well as an evaluation of the methods investigated. Besides, this section offers a detailed outline of the algorithm's development and its implementation. The series of algorithms presented in this research project are all part of the design vocabulary presented as the practical outcome of this thesis, which is proposed to help other artists and designers to create their own aesthetic experiences.

A large portion of this research project's attention and effort was devoted into the investigation and creation of a sound engine, which was anticipated to implement both the theoretical and practical lessons learned during the development of the collaborative projects. This thesis aims to make a logical, functional and sensible use of sound in the creation of audiovisual aesthetic experiences. A key motivation of this research project is to translate sound into motion, that is, to generate motion from the information extracted from the audio file. The idea of extracting information from audio data and translate it into motion was created during my work in the collaborative project *Surround Sounds* with Christian Marclay, as outlined in chapter 4.5. My mission as audiovisual designer in the project was to come up with a vocabulary of motions for the onomatopoeias collected by Marclay.

The concept of *Surround Sounds* was to translate sound into motion. So, I realised that sounds of objects and events could be recorded and have their waveform analysed, evaluated and employed to the creation of motion. Following, after a period of testing, I implemented an existent manually achievable method for extracting data from audio files and apply it to change the position and rotation of the onomatopoeias in real-time, syncing audio and visual material. This technique acknowledges a wide range of different relationships between audition and vision as explored in the chapter 2.5 of this thesis. Moreover, it contributes in practical terms to the debate regarding the congruence between the sonic and the visual domain that was initiated in section 2.5 of the literature review.

AudioVisual Collage sound engine performs FFT analysis to detect the highest amplitude frequency. That is, the frequency that contains the highest level of energy, in terms of amplitude, is automatically detected and applied to the generation of visual motion. The following section of this chapter will describe the implementation of this process in detail.

8.2.1. Extracting Data from Audio Files

The AudioVisual Collage sound engine was implemented using the Minim audio library in Processing. Minim is an open-source library that extends the capabilities of Processing and includes advanced audio options, which was originally created by Damien Di Fede and further developed by Fede, Mills and Brix (2010). The Minim library is being currently distributed as part of the standard installation of Processing, as one of its core libraries.

In order to allow the AudioVisual Collage system to access the files for processing, the AudioVisual Collage software prompts the participants to upload images as well as audio files, which can consist of mono or stereo tracks of several audio digital formats, including: of WAV, AIFF, AU, SND, and MP3. Next, the platform extracts data from the file, which resides outside the programme folder. However, before it can be used, it needs to be loaded into the programme. That is, the file needs to be loaded into the buffer so the software can extract the necessary information and start drawing the program. To achieve that, the platform prompt the user to indicate the location of the file desired to be loaded into the program, by using the `selectinput()` function (Figure 8.5).

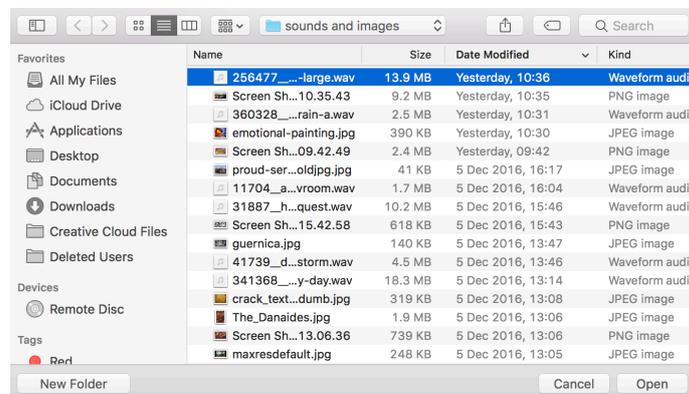


Figure 8.5. Dialog box where participants introduce images and audio files.

AudioVisual Collage opens a dialog box that offers the participants the opportunity to select an audio file for input. After the selection is made, the programme gets the location of the file using `getAbsolutePath()`, which is a method that detects the directory where the file is located and returns the absolute path name string of the file selected, which is then input into the audio player (Figure 8.6).

Code:

```
void fileSelected(File selection)
{
  if (selection == null) {
    println("Window was closed or the user pressed cancel.");
    exit();
  } else {
    minim = new Minim(this);
    song = minim.loadFile(selection.getAbsolutePath());
  }
}
```

Figure 8.6. Introducing an audio file selected by the viewer into the processing chain.

The first attempt to implement a method for generating visual motion from audio data was done by splitting the sonic spectrum into three frequency ranges, which were: 20Hz to 150Hz; 151Hz to 2.5KHz; 2.6KHz to 20KHz, and later, to assign the information extracted from each of the three frequency ranges to each of the three generated from the image segmentation process in order to individually generate motion to the layers.

In order to fragmented the frequency range into three discrete frequency bands a Fast Fourier Transform [FFT] algorithm was used in order to generate a sonic spectrum, which was then divided into average bands using `linAverages()`. To explain, the FFT algorithm decomposes the audio signal into partial waves so as to make it more easily analysable (Heideman et al., 1985).

The FFT values can be understood as the amount of any specific frequency that is present at a specific moment in time. To explain, for every segment of samples that are used to calculate an FFT, the FFT produces a series of values that represent the amount of energy across a range of frequency bands. With respect to the Real Fast Fourier Transform [RFFT], the number of frequency bands is half the size of the segment of samples utilised as input (Heideman et al., 1985). Therefore, if 2048 samples are used as input,

1024 RFFT values are output. Each of these outputs contains a level of energy that indicates how loud each of the frequency ranges are in a series of bands (bins). The width of the frequency bands can be calculated by dividing the audio file's sampling rate by the FFT size. So, with a sampling rate of 44100 and an FFT size of 2048, the width of each band (also known as a 'bin') is: $44100/2048 = 21\text{Hz}$ approximately. So, the first of the 1024 (FFT size/2) values contains energy between 0 and 21Hz, the second contains energy between 21Hz and 42Hz, the third contains energy between 42Hz and 63 Hz, and so on (Heideman et al., 1985).

To explain, the FFT class analyses an audio stream and fills an array with samples in the frequency domain (bins), therefore, it allows the use of different frequency range bands of a particular audio file to individually trigger multiple events and to visualise them. After this FFT analysis was completed, a Beat Detect algorithm was implemented so as to detect sonic impulses in different frequency ranges, for instance low, mid and high frequency bands, and to assign them to the separate visual segmented layers.

However, the approach of separating the sonic spectrum into three frequency ranges and applying the information extracted from each of the frequency ranges to the movement of the three separated image layers presented a major drawback. That is, when analysing the digital audio recording of a distant rumble sound (Figure 8.7), it presented a very low level of high frequency content.



Figure 8.7. Audio recording of a Distant Rumble sound
 (A) Low frequency range content
 (B) Middle range frequency content
 (C) High frequency range (no information detected)

Thus, the layers which had their movements triggered by the sonic content present in the high frequency range would not move. Consequently, the resulting motion would be inconsistent. That is, the system would rely on the amount of content present in each of the frequencies. As a result, depending on the sound input by participants one or more layers could potentially remain static throughout the audiovisual aesthetic experience.

The solution for this issue was the implementation of a method to detect the frequency that contained the highest amplitude. The technique consists of performing FFT analysis on the audio file so as to detect the highest amplitude frequency within the frequency spectrum and to address this portion of the sound file to the trigger the motion of the visual content (Figure 8.8).



Figure 8.8. Audio recording with highest amplitude frequency of 240Hz.

As a result, AudioVisual Collage extracts information so as to generate consistent movement independently of the frequency response of the audio files input by participants. This technique is described in detail in the following section.

8.2.2. Highest Amplitude Frequency Detect

The ability to control an object's motion using information extracted from an audio file was a technique that I have employed in the development of the collaborative process with Christian Marclay, which was presented in

chapter 4. In *Surround sounds*, which consisted of a silent piece, this technique was used to generate templates of movements that were later applied to all the individual onomatopoeias extracted from comic books. This technique implemented during the collaborative project with Marclay served to inform the implementation of this algorithm, which is now part of the AudioVisual Collage sound engine, and automatically completes the process of applying motion into visual elements.

The technique applied in the development of *Surround Sounds* was manually controlled. That is, particular sounds were selected and had their sonic content analysed and turned into keyframes, as described in chapter 4.3.2. However, in AudioVisual Collage, the participants are prompted to input sound files that can drastically vary in terms of level and frequency response. Therefore, it was required to implement a system capable of consistently certifying that a minimal amount of data would always be extracted from the sound files and applied to the creation visual motion.

As outlined in section 8.2.1 of this chapter, an FFT object was used to convert an audio signal into its frequency domain representation. This frequency spectrum representation allows the system to analyse the amount of energy present in each individual frequency of the audio signal. Then, in order to certify a constant motion, the algorithms extracts data from the frequency that containing the highest amplitude within the frequency spectrum (Figure 8.9).

Code:

```
// go through the frequency ranges to identify the one that has the highest amplitude
for (int n = 0; n < fftLin.specSize(); n++) {
    if (fftLin.getBand(n) > fftLin.getBand(highest))
        highest=n;
}
```

Figure 8.9. Algorithm that identifies the audio frequency range with highest amplitude.

8.2.3. Generating Motion from Audio Data

Once the highest amplitude frequency was detected, data was extracted from this particular frequency over time and applied to the generation of motion. For instance, the audio recording of a heavy rain (Figure 8.10) presented a higher intensity on the high frequency range of 5KHz.



Figure 8.10. Frequency response of the audio recording of a heavy rain.

Thus, the algorithm recognised this specific frequency range, extracts data from it and applies that data to transform the z-axis position of the front layer using the `translate()` function.

Code:

```
//extracts it from the sound and applyies to
the z-axis position of the image's front layer
pushMatrix();
translate(0,0,highest);
beginShape();
tint(255,225);
texture(img1);
vertex(0, 0, 0, 0, 0);
vertex(709, 0, 0, img1.width, 0 );
vertex(709, 443, 0, img1.width, img1.height);
vertex(0, 443, 0, 0, img1.height);
endShape();
popMatrix();
```

Figure 8.11. Information extracted from the highest amplitude frequency of an audio file range being used to control the z-axis position of a quadrilateral shape.

Later, the sum of all frequencies, which formed the full spectrum of the sound, was applied to the middle ground layer. Thus, both layers move synchronously, yet present variations since, in this case, the content extracted

from the low frequency range was also applied to the middle layer. As the image below illustrates (Figure 8.12), the waveform that represents the frequency with highest amplitude presents more defined transients and peaks. As it consists of a single frequency, it presents more separation between the lowest and highest areas. As a result, each of the waveforms provides data that will be applied to the generation of motion. The resulting motion will have complete synchrony but still present variations.

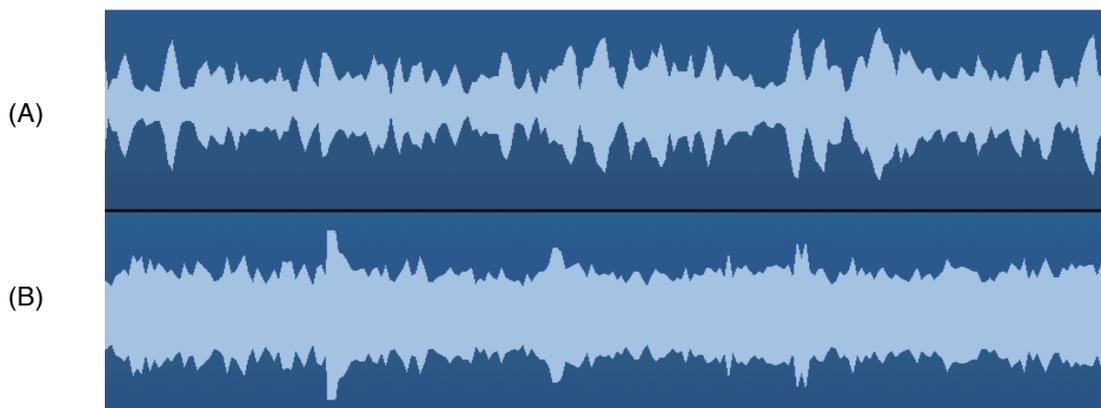


Figure 8.12. (A) Highest amplitude frequency in the Heavy Rain sound recording
(B) Full spectrum waveform of the Heavy Rain recording

The image below displays another example of a sound that presents higher intensity in a specific frequency range (Figure 8.13). The sound recording of a single note being played in a French horn presents higher intensity in the mid-range between 450hz to 2KHz.



Figure 8.13. Frequency response of a French horn single note.

The highest amplitude frequency detect algorithm identifies a particular frequency within this range and applies it to z-axis of the visual layers so as to generate motion within 3D space. In the same way that it was performed previously example, the sum of all frequencies, which formed the full

spectrum, is applied to the middle layer, so both layers present variations while moving in synchrony.

The approach presented in this section automates a technique used in the *Surround Sounds* collaborative process. Hence, the AudioVisual Collage sound engine algorithm consists of an efficient method for automating a particular creative decision. This method is part of the process of translating sound into motion, which was pursued during the collaborative creative process with Christian Marclay. This approach emulates the way in which an object would vibrate and resonate when exposed to a sound stimulus; specially if that sound stimulus happened to vibrate in the same resonating frequency as the object. Thus, it aims to mimic a process present in the real world so as to deal in a holistic way with a complex processes, that involves experience, perception and understanding.

The algorithm presented in this section was implemented as part of the AudioVisual Collage audio engine. Nevertheless, it comprises part of the design vocabulary for audiovisual composition proposed by this thesis. The proposed algorithm applies a well known technique, which consists of detecting the frequency that contains the highest amplitude in an audio signal, which was achieved with the implementation of FFT analysis. However, the information extracted from this process is implemented in a new way. That is, the information extracted from the highest amplitude frequency over time was processed, turned into usable data and applied to the transformation of the position parameters of the visual elements.

This chapter presented the development of AudioVisual Collage and described how the essential techniques and design principles identified during the collaborative process, outlined in chapter 4 and 5, were incorporated into the final project. This chapter also revealed how these techniques were turned into algorithms that together, as a series of automated steps, form the Audiovisual Collage software.

8.3. AudioVisual Collage Output

(See the accompanying folder 'AudioVisual Collage Project' for all video excerpts, images, audio files and AudioVisual Collage Software).

This section will outline four outputs generated by AudioVisual Collage. The series of outputs are presented to display the resulting audiovisual aesthetic experience created by the system using images of paintings and digital graphic design (Table 8.1). The series of outputs are organised as videos excerpts that can be accessed in the subfolder 'AVC Output' located inside the accompanying folder 'AudioVisual Collage Project'. The final system can be accessed in the subfolder 'AudioVisual Collage Software'. The software can be operated as a Processing sketch, which contains the code and all libraries used, or as a standalone application available for OSX and Windows platforms.

Car.mp4	Car_painting.jpg	Car_engine.wav
Bombings.mp4	Guernica_painting.jpg	Warfare_ambience.wav
Pop.mp4	Pop_Onomatopoeia.jpg	Bubbles_popping.wav
Sea_of_Fog.mp4	Wanderer_Above_the_Sea_of_Fog_painting.jpg	String_wind.wav

Table 8.1. List of outputs generated by Audiovisual Collage software.

The video excerpt entitled 'Car.mp4' demonstrates the first output presented in this section. The files input into the system so as to generate this audiovisual output are the oil painting of a car and the audio recording of a car engine. The image below illustrates the complete visual process performed by the image segmentation and inpainting algorithm (Figure 8.14).

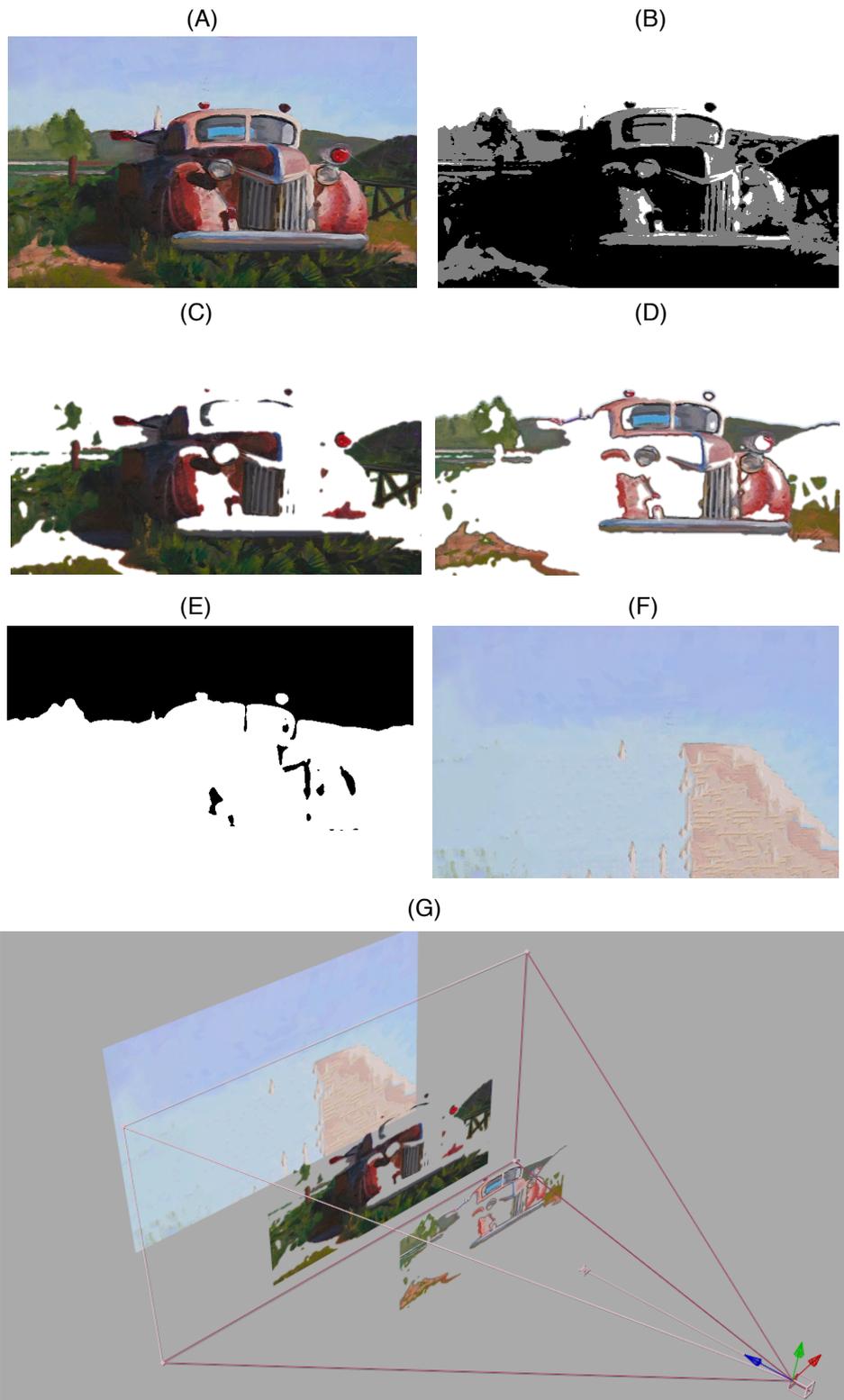


Figure 8.14. AudioVisual Collage visual processing.

- (A) Original image with brightness mean value of 132.3
- (B) Image Posterization process
- (C) Middle layer segmented
- (D) Front layer segmented
- (E) Alpha mask to map inpainting area
- (F) Background Inpainting
- (G) 3D Composition

As outlined in chapter 6.6, the AVC Image segmentation and inpainting algorithm goes through all pixels contained in the image and calculates the intensity of each one, in order to provide the image's brightness mean value. In this instance, the image's brightness mean value consists of 132.2, which is more than the global average defined as 127. Hence, the system classifies the lighter portion of the image as background and automatically applies inpainting to it. Accordingly, the proposed algorithm successfully performs the task of automating a creative decision as it emulates an important artistic technique I employed throughout the collaborative project Interactive Art Gallery, as outlined in chapter 5.5.

As the video excerpt 'Car.mp4' demonstrates, the image segmentation and inpainting methods have presented effective results as the image was correctly segmented into three discrete layers and applied as texture to quadrilateral shapes within 3D space. In addition, the 3D lighting effect, which is controlled by the image's brightness mean value, plays an important role in maintaining the sense of depth and separation between the layers within 3D space.

As the image below reveals (Figure 8.15), the 'Car_engine.wav' audio recording contains higher amplitude in the frequency of 750Hz.



Figure 8.15. The frequency response of the sound recording of a car engine shows the 750Hz as the highest amplitude frequency.

The AVC sound engine algorithm performs FFT analysis so as to detect the frequency that contains the highest amplitude and extracts its data over time. Next, once data is extracted, it is applied to the z-axis of the front layer in order to generate visual motion with the use of the translate() function.

Using the same process, data is extracted from the sum of all frequencies (full spectrum) and applied to the middle layer using the same approach. The resulting audiovisual aesthetic experience exemplify, in practical terms, a method for manipulating the temporal structure and spatial organisation of 2D visual elements so as to reflect their content within 3D space. What is more, the system presents an approach for combining visual artworks and sound in the creation of algorithmic screen-based aesthetic experiences.

(A)



(B)



Figure 8.16. (A) Waveform of the sound recording of a warfare.
(B) Separate layers movements are triggered by the audio file.

The second output of the series, entitled 'Bombings.mp4', displays a different approach. This example was generated using the painting *Guernica* (Picasso, 1937) and the audio recording of warfare as inputs (Figure 8.16). The painting *Guernica* consists of a cubist representation of the bombings in the Spanish town of Guernica, and the sounds of explosion and shootings were used so as to generate motion in the segmented layers of the image. The warfare audio recording has 275Hz as its higher amplitude frequency. Thus, data extracted from the audio files, presented short bursts of low

frequency followed by high frequency hissing sounds. The resulting motion resembled the vibrations caused from a bomb explosion, which were generated from data extracted from the low frequency range of the audio recording.

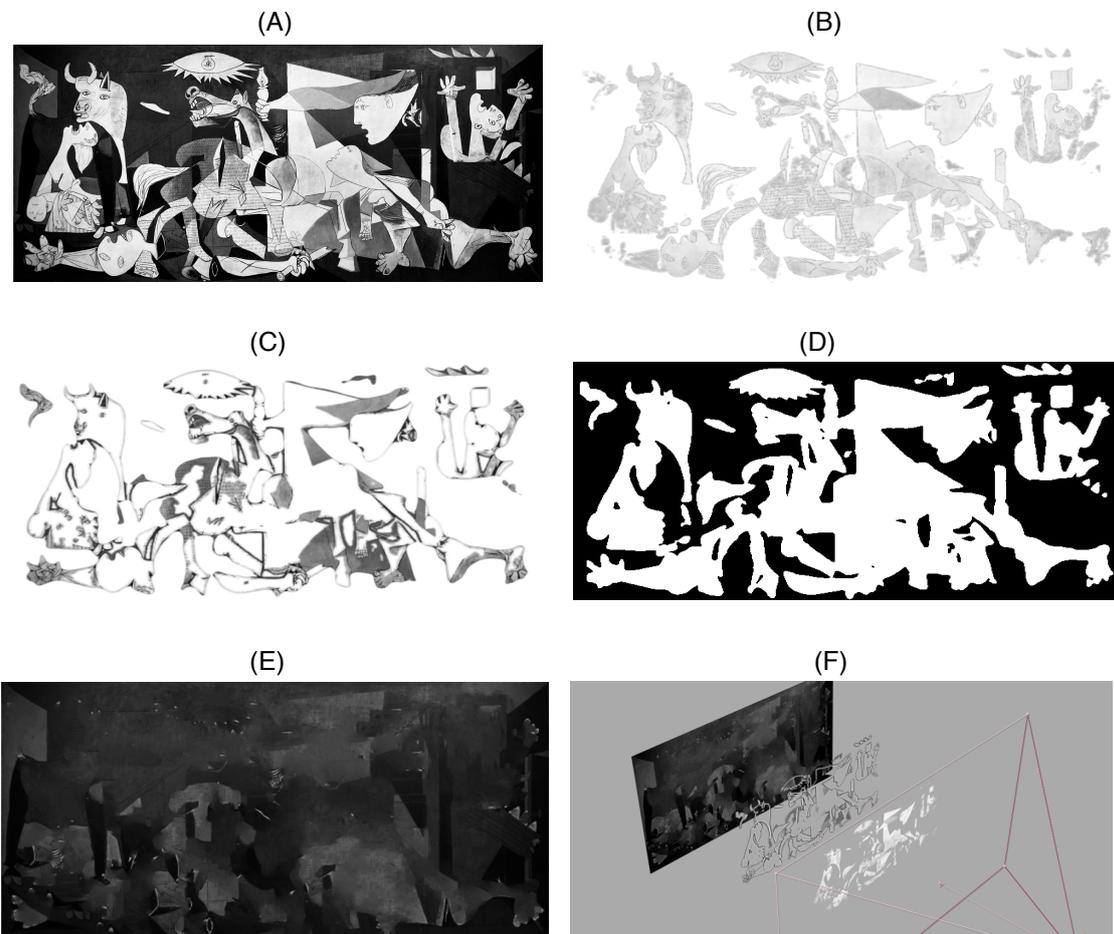


Figure 8.17. AudioVisual Collage visual processing.

- (A) Guernica painting, with brightness mean value of 101.6
- (B) Front layer segmented
- (C) Middle layer segmented
- (D) Alpha mask to map inpainting area
- (E) Background Inpainting
- (F) 3D Composition

The digitised version of the painting *Guernica* presents the brightness mean value of 101.6. Thus, the system classified the darker portion of the image as background and automatically applied the inpainting method to it. As the image above shows (Figure 8.17), the edge of the elements segmented and classified as part of the front layer, were separated and categorised as

middle layer. The same effect occurred in the edges of the middle layer, causing the background layer to present lighter grey pixels, which were introduced during the inpainting process. Yet, once the layers are placed within 3D space and overlapping each other, the undesired effect partially disappears, because the inpainted portion of the image is hidden behind the other layers, as the video excerpt 'Bombings.mp4' demonstrates. Nevertheless, this comprises a topic that could be further analysed and addressed in future iterations of the project.



Figure 8.18. AudioVisual Collage visual processing of Pop.mp4
 (A) Pop image, with brightness mean value of 114.2
 (B) Image Posterization process
 (C) Front layer segmented
 (D) Middle layer segmented
 (E) Background Inpainting
 (F) 3D Composition

The third output of the series consists of a graphic illustration that contains the onomatopoeia *Pop!* (Figure 8.18). The process of image segmentation and inpainting presented an effective result. As a matter of fact, this output is a good example of the how efficiently the AVC image segmentation and inpainting algorithm performs when processing digital graphic art that are based on the use of lines and tones.

The video excerpt entitled 'Pop.mp4', shows the system generating the output in the same context in which it was first envisaged, during the collaboration with Christian Marclay in the piece *Surround Sounds*. To explain, in order to generate this output, AudioVisual Collage extracts data from the highest amplitude frequency present in the audio recording of a sequence of bursting bubbles. Then, it applies this data to transform the position of the front layer's z-axis, using the `translate()` function. The system, in this context, generates movements that resemble the way in which an object producing this specific sound stimulus would behave, in terms of motion.

The fourth output of the series, named 'Sea_of_Fog.mp4', was generated using the painting *Wanderer Above the Sea of Fog* (Friedrich, 1818), and the audio recording of a strong wind (Figure 8.19) as input files.



Figure 8.19. Frequency response of the audio recording of the wind.

The painting shows a person contemplating a landscape that reveals the presence of fog and possible wind (Figure 8.20). Thus, the AVC sound engine algorithm detected the highest amplitude frequency, which is in this case 135Hz, and extracted data from the audio recording so as to generate motion that resemble the way objects are affected when subject to the forces of the wind.

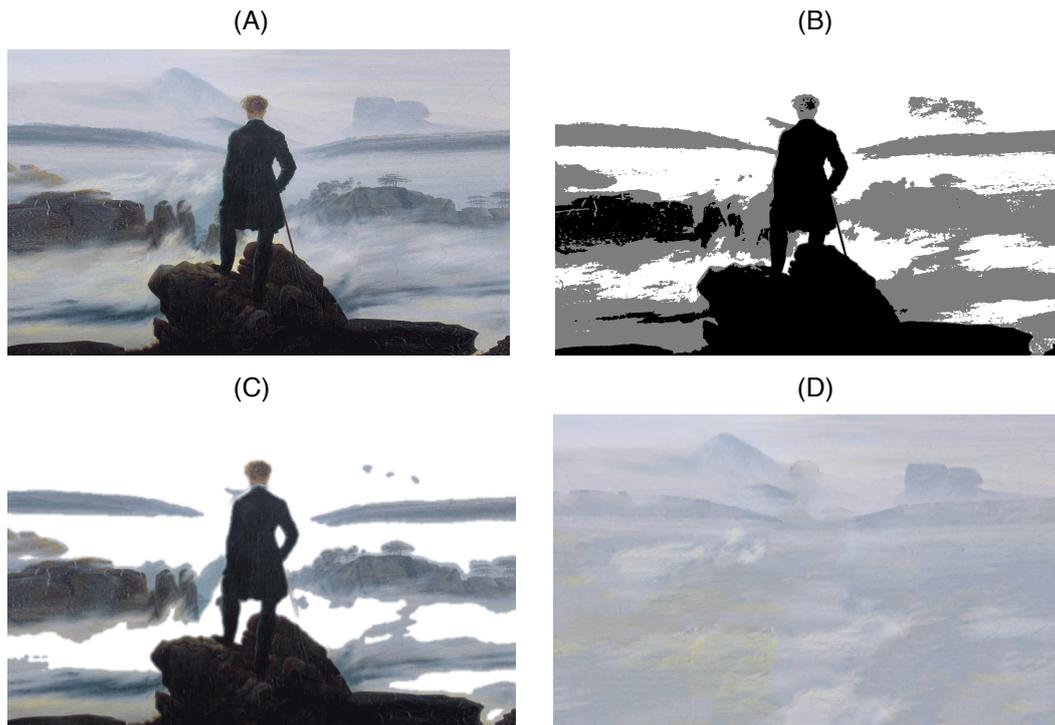


Figure 8.20. AudioVisual Collage visual processing of Sea_of_Fog.mp4
 (A) Wanderer Above the Sea of Fog painting (brightness mean value 151.6)
 (B) Image Posterization process
 (C) Image Segmentation method
 (D) Background Inpainting

This section has revealed outputs generated by AudioVisual Collage in different contexts: digitised versions of paintings and graphic illustrations. Under those circumstances, the AVC image segmentation and inpainting algorithm presented effective results manipulating the spatial organisation of 2D images so as to reflect their content within 3D space. Equally important, the AVC sound engine algorithm proposes an original way of combining visual artworks and sound for the creation of algorithmic screen-based audiovisual aesthetic experiences.

AudioVisual Collage successfully automates key steps of my creative process identified during the development of the two case studies presented in chapter 4 and 5 of this thesis. The contribution of AudioVisual Collage to the computational art field is twofold. Firstly, it proposes a method that combines completely automated image segmentation, inpainting and 3D composition. Secondly, the system proposes a method for generating visual motion using data extracted from audio files.

As outlined in chapter one, the practical outcome of this research project consists of the materialisation of an artistic vision in the form of a platform for art practice. AudioVisual Collage supports the creation of artworks by allowing the participants to input their own audio, and visual material, so as to create personal audiovisual aesthetic experiences. Therefore, AudioVisual Collage consists of a series of algorithms that offer great creative potential.

Chapter 9. Conclusion

“When, by group action, a society evolves a new medium, it has earned the right to express a new message”.

(Marshall McLuhan, 1959)

9.1. Summary

This research project was developed with two main purposes. Firstly, to produce a series of automated tools that emulate the essential artistic techniques I employed throughout the collaborative process presented in chapters 4 and 5. These automated tools were refined, combined and implemented into the *AudioVisual Collage* system, which was presented in chapter 6. Secondly, to generate a design vocabulary for screen-based audiovisual composition defining the technology and technical skills required to build and to implement these automated tools. An iterative creative process as well as an investigation into the field of computational art supported the development of this project.

This thesis was completed with the combination of research and practice. The creative process completed in collaboration with other artists and creative groups was the main vehicle to identify my artistic approaches and techniques. The theoretical investigation revealed important opportunities for enriching both, the collaborative process and the creation of automated tools for screen-based audiovisual computational art practice.

A survey of the currently available software and techniques provided a strong foundation that helped me identifying what technology currently available could be used for the development of this practice-based research. This thesis explores both sonic and visual fields of study so as to explore their relationship, which was vital for the creation of an innovative creative system. The outcome of this research has brought me a new perspective in relation to audiovisual art practice. The outcome of this project consists of a design vocabulary in the form of a set of algorithms designed to perform a series of

automated processing tasks.

- Automatic Image Segmentation
- Automatic Inpainting
- Automatic 3D Composition
- Automatically Controlling 3D Lights with image brightness
- Automatically Generating Visual Motion from Audio Data

The review of the literature has investigated the configurations of the human and the machine in the field of computational art. It has also explored the concept of *aesthetics* and *perception* in order to look beyond the technological and to contemplate the creative and experiential side of the digital domain. The research into aesthetics and perception, within the context of computational art, has been highly relevant to the development of AudioVisual Collage. That is, when one contemplates the degree to which images and sounds are embedded features of our environment, it is vital to understand the way in which we perceive audiovisual computational art.

This thesis has been strongly influenced by Chion's notion of synchresis, as outlined in chapter 2.5, in which the rapid creation of a relationship is produced between sonic and visual information when observed in synchrony. This project also follows the audiovisual composition approach (Grierson, 2005; Correia, 2015), which considers visual and sonic events as one single perceptual unit. Besides, this concept has been explored and applied in practical terms, with the creation of an algorithm that extracts data from audio files and converts it into visual motion.

9.2. Discussion

9.2.1. AVC Image Segmentation and Inpainting Algorithm

The AVC image segmentation and inpainting algorithm consists of a combination of two techniques identified during the development of the collaborative project Interactive Art Gallery, presented in chapter 5. These techniques consist of important creative steps that I applied as audiovisual practitioner engaged in a collaborative process with other artists. These techniques were implemented with the use of tools that require a considerable amount of interaction and expertise to be accomplished. For this reason, the purpose of this algorithm was to propose a method to automate these tasks so as to allow other artists and designers to quickly achieve similar results.

It must be remembered that during the development of Interactive Art Gallery, the artists involved in the creative process were asked to pinpoint important elements in their artworks. Then, the image segmentation and inpainting techniques were manually applied in order to extract the elements and fill-in the gaps left in the background image.

However, while operating AudioVisual Collage, participants are requested to input images and audio files. The image files input can potentially be of any type, from representational images to completely abstract ones. Thus, the major challenge encountered during the development of the AVC image segmentation and inpainting algorithm was to create specific rules and constraints to certify that any input image would be successfully segmented and have its background inpainted in a coherent and consistent way. Besides, the algorithm was required to identify which of the resulting segmented layers consisted of the background.

As outlined in chapter 6.4.2, the solution for this issue was the implementation of a posterization method, which resulted very effective for two main reasons. Firstly, the posterization method certifies that all images, of any type, are always automatically segmented into three discrete layers and placed within 3D space. Secondly, the posterization method reduces the

number of colours to only three tones of grey. Thus, it considerably increases the algorithm's performance, in terms of execution time, and allows the image segmentation and inpainting tasks to be completed much faster.

As the chapter 6 describes, the implementation of the inpainting method presented two fundamental challenges. The first challenge was to design a rule to automatically define which of the layers generated by the image segmentation process consisted of the background to be inpainted. The second challenge was to automatically define which area of the background image required inpainting. To clarify, the inpainting technique used in the development of Interactive Art Gallery was performed using the clone brush tool, found in Photoshop CS6, which is a well known approach to inpainting. However, similarly to the image segmentation method used in the collaborative process, it requires a certain level of expertise and a great amount of interaction to be accomplished, as it involves manually copying nearby pixels into the gaps, using the clone brush tool.

The solution for the first challenge, as described in chapter 6.4, was to calculate the image's brightness mean value and use this information to define which portion of the image consisted of the background, middle ground and foreground, which presented stable results. The solution proposed for the second challenge, which consisted of defining the area to be inpainted, was also effective. The solution, as revealed in chapter 6.4, was combining the area of the foreground and middle ground layers generated during the image segmentation process for the creation of an alpha mask. Then, the mask is used as a map that informs the inpainting algorithm where to look for nearby pixels and in which region to place them.

The proposed algorithm combines both image segmentation and inpainting techniques and, as the outputs demonstrate in chapter 8, consists of an effective tool for automatically extracting elements from 2D images and manipulating them so as to reflect their content within 3D space.

9.2.2. AVC Sound Engine Algorithm

The essential design technique implemented during the development of the collaborative project *Surround Sounds*, was the extraction of data from audio files for the generation of visual motion. As outlined in chapter 4, this process involved isolating the more representative frequency range of a specific sound. This approach was performed manually, which resulted a very time consuming and laborious process. The process was done with the use of a digital audio equaliser to increase the most representative frequencies of a given sound before converting it into keyframes, in order to be used to the generation of visual motion.

As outlined in chapter 8.2, the AudioVisual Collage sound engine algorithm was designed to automatically perform this task. The proposed algorithm applies FFT analysis to detect the highest amplitude frequency. Then, it extracts data from that specific frequency over time and applies it to the position parameters of the visual elements to generate motion.

The implementation of this technique, which is the core of AudioVisual Collage sound engine, has presented successful results. As the outputs presented in chapter 8.3 show, the algorithm consistently generates sufficient data over time. Thus, it ensures that the position parameters of the visual elements will present changes, in response to the sounds input into the system. To put it simply, the underlying idea is to certify that a substantial amount of data is extracted from the audio file and applied to the generation of visual motion.

Also, when compared with the Beat Detect method, which was previously implemented into the project during the early stages of development, it presents better results. The highest amplitude frequency detect method is considerably superior because, as the results presented in chapter 8.3 demonstrate, it produces a closer association, in terms of motion and synchrony, between the audiovisual input and the obtained output.

Another important aspect that contributes to the success of the proposed algorithm is that the resonant frequency of objects often carries a

large amount of sonic energy. Thus, this method mimics a phenomenon present in the real world. That is, by generating motion on visual elements, using data extracted from the highest amplitude frequency of audio recordings, AudioVisual Collage generates an audiovisual relationship that emulates the way in which a specific object would behave and react, in terms of vibrations and motion, when a sound stimulus is produced.

This research project explores the affordances of the digital domain to generate new proprieties that were non-accessible in analogue traditional media. This reveals the real *essence* of computational art, which consists of a continuous and stimulating adaptation of existing techniques for the construction of new ones.

9.3. Art as Research

An important lesson learned during the development of this project, is that, for an environment to successfully foster creativity, it demands not only artistic expertise and appropriate technical facilities, but also, the ability to support the elaboration of *new* approaches, which in turn, empower the gain of new knowledge, skill and experience.

The main concern when defining the methodology of this project was how to guarantee that both the technological and artistic development was equally nurtured alongside each other. To assure the artistic and technical aspects were equivalently fostered this project made all techniques created to support the aesthetic development of this project was highly responsive and part of an iterative process. Therefore, the insights were re-introduced into the creative process of development. Nevertheless, the user study has generated a great amount of feedback, which could only partially be re-introduced into the project, but has undeniably been an important guide to the development of this research.

As a result, this practice-based research comprises a co-evolutionary process, where new technologies were invented and existent technologies

were adapted in a new ways. Thus, it supported the transformation of existent artistic forms and traditional practices into novel ones.

Collaborating in projects where I was requested to create artworks, allowed me to obtain an art practitioner's perspective. For that reason, the approach taken in the development of this thesis was to investigate the work in progress, producing a user study and generating reports of its results along the way, in order to complement the retrospective descriptions of the creative process and its outcomes. This approach led to the conception of a structural design vocabulary produced to facilitate and investigate creative practice, and to support other artists and designer in both theoretical and practical ways.

The user study introduced to the participants, *AVC the Image Segmentation and Inpainting* prototype, presenting two algorithms combined in an image segmentation and inpainting tool. The study outlines that special attention should be paid in order to reduce the trade-off that results from incorporating content and new functionalities, and improving the explorability and easy of use. As programmer Joshua Noble (2012) advises:

“Although engineering is often the art of avoiding failure, interaction design is the art of failing and understanding the failures correctly”.

The study provided an efficient evaluation method to be implemented in projects in the field of computational art. Moreover, the user study suggests paths for future improvements in screen-based computational artworks, which is based on the suggestions and insights given by the professional artists and audiovisual designers that took part in the study.

9.4. The Art Studio as Laboratory

During my PhD studies, I have taken an active part in the development of projects as researcher actively engaged in collaborative art projects. This approach allowed me to acquire tacit knowledge and experience that directly emerge from both the investigation and the art practice. In doing so, I had the

opportunity to take part in creative collaborative processes with other artists so as to assist them to convey their artistic messages in the digital domain. This approach demanded a high level of commitment and involvement as all projects developed during this project were internationally published, in art galleries and online platforms, in the United Kingdom, France, Switzerland and the United States. Besides, since the collaborative projects comprised tangible artworks implemented in real-life situations, they helped me to generate a perspective inaccessible for an outside observer.

During the development of the *Interactive Art gallery* project, described in chapter 5, I was asked to develop an interactive system to be embedded in the Debut Contemporary Art gallery website, to generate attention and promote artists from the gallery by exploring their artworks in the digital domain. For that reason, I had to communicate with other artists so as to recommend new ways of exploring their artworks in the digital domain. Therefore, the projects created during my PhD studies originated from first hand experiences and experiments with novel design techniques.

In the Surround Sounds project, outlined in chapter 4, while working as part of the creative team of Christian Marclay, I had the opportunity to both learn and, also, to apply my competence and experience in order to contribute to a successful outcome. However, my main concern has always been reflecting on my own art practice, so this self-evaluation would become an instrument to help me developing an understanding that would be inaccessible from an outside viewpoint.

The collaborative process revealed that most challenging aspect of creating an environment to facilitate collaborations between artists, scientists, and researchers, which is to effectively encourage the exploration of new territories and to stimulate the sort of communication that leads to an effective exchange of perspectives and knowledge. To summarise, my approach as an art practitioner involved in a collaborative process allowed me to be actively and deeply involved in every stage of the projects development, while at the same time, evaluation the creative process with my research questions in mind.

The introductory chapter of this thesis addresses its audiences, which are: visual artists, audiovisual designers and people with interest in conveying their artistic messages in the digital domain. Therefore, at this conclusive stage, it seems an appropriate moment to revisit this original purpose and outline what specifically are the contributions of this research project for these groups.

The algorithms that compose the AudioVisual Collage software were originated from the analysis and evaluation of the collaborative projects. That is, the main techniques utilised to create the collaborative artworks were identified and later turned into algorithms, to allow other artists to create their own artworks. Throughout this process, this research has presented and developed a number of concepts to guide and facilitate the creation of screen-based aesthetic experiences in the digital domain. As first explained in the introduction chapter, and then repeated on multiple instances through this thesis, the process of development of this research project has been an iterative process, completed by creating, drawing on, discarding, combining and adapting concepts.

As shown above, this thesis contributes to the field of computational art with the creation of models that have been developed, introduced and discussed. The methods, techniques and algorithms presented in this thesis comprise a combination of novel inventions and an iterative refinement of pre-existing technology and practice. Therefore, these approaches and techniques are expected to have importance and value to audiovisual artists as they provide a design vocabulary for art practice in the digital domain.

As has been noted, this project incorporates approaches from traditional media, such as sound recording and cinema that for many mechanical reasons are linear. Nevertheless, this research project collects those techniques and re-arranges them in a non-linear fashion. The result is a hybrid and rich audiovisual media, which is composed of a mixture of numerous artistic languages.

In essence, AudioVisual Collage software aims at finding an exciting balance between the artists' control and lack of control, over the aesthetic

experiences created using the system. AudioVisual collage generates content that is dynamically constructed and updated in real-time. Therefore, by creating artworks using the system, the participant engages with dynamic outputs of computation that occurs instantaneously. Although static data might be involved, as the user inputs static files containing images and audio content, the AudioVisual Collage platform cannot be reduced to a single document stored in a specific type of media. That is to say, in contrast to works of literature, music scores, paintings or films, there is no possible way to consult a single file that could retain all the possible outcomes that can be generated by AudioVisual Collage. The research project proposes a new tool for art practice in the field of computational art, and gives artists flexibility and power to create their own aesthetic experiences.

As section 5 of the review of the literature revealed (Murray, 2011; Manovich, 2002), software is increasingly becoming more democratic and is, therefore shaping our culture. AudioVisual Collage is part of this new type of media, which is now a fundamental feature of all areas of society. Thus, it is vital to consider digital systems in order to understand contemporary culture and society. AudioVisual Collage aims to empower artists to express their artistic ideas within the digital domain. Thus, it is also part of this process of understand how our culture is being shaped by digital systems and how these digital system, in turn, are shaping our culture.

Viewing from a technical perspective, AudioVisual Collage software comprises a series of algorithms that emulate and automate my artistic approaches, through control statements such as while/do and if/then, classes, data types, data structures and dialogue boxes. However, when looked from an artistic point of view, AudioVisual Collage consists of an invisible glue that ties an entire creative process together. The proposed system expects to be utilised, explored and developed further by other artists. Therefore, entirely new ways of creating art in the digital domain are expected to be invented. The types of artworks that can potentially be created using AudioVisual Collage software are vast. That is, visual and audible content input into the system can be chosen so as to orchestrate the aesthetic journeys in distinct

ways. Therefore, the art pieces are automatically generated by a series of algorithms and at the same time, actively constructed and managed by the participants, based on their choice of content. On balance, this research project expects to contribute to the narrowing of the art and computational technology divide.

Altogether, this research project expects other artists and designers to find the design ideas embodiment by the prototypes interesting and inspiring. Even though the two prototypes presented have been developed as part of this research project, they are expected to have a life of their own that goes past the limits of this thesis. As the next section of this chapter outlines, the first series of artworks created using AudioVisual Collage software has already been produced and is currently being commercialised. Thus, the proposed software has been exposed to a wide audience.

In fact, the AudioVisual Collage software has already received a considerable amount of media attention. Therefore, since the platform created during this research project is potentially going to reach a much larger audience than this thesis, the prototype was designed to embody, stand up for, and speak of the arguments that have been made throughout this thesis. After all, the practical portion of this this research project can be seen as a concrete demonstration of the arguments and concepts presented in this thesis.

9.5. Future Developments

This research offers many opportunities for future development. Some important future development suggested during the user-study, were the ability to share the content created, and also, for collaboration. In the present version of AudioVisual Collage the participant are requested to input an image and a sound file. Nevertheless, an idea for future development would be allowing multiple users to simultaneously input images and audio files in order to create aesthetic experiences in collaboration.

Also, there is a constant issue with regards to computer power. That is, all the artworks generated by AudioVisual Collage are created in real-time. That is, multiple algorithms applied complex and expensive, in terms of computer power, processes in order to create the audiovisual aesthetic experiences. It currently takes four seconds to generate an audiovisual aesthetic experience running at 60fps, which derives from a digitised image and a three minutes long *mp3* stereo audio file. This is the time taken to play a *360p* video, plus audio, in a MacBook Pro Retina, Processor 2.5 GHz Intel Core i7 with 16 GB of RAM memory.

It would be a great advance to be able to generate higher video output, as it would facilitate the creation of high definition 1920x1290p video files at 60fps, and allow the real-time rendering to have a great improvement in both its stability as well as image quality, particularly for the inpainting process. Because, having images in a greater number of pixels, would allow the inpainting algorithm to create even more details in the process of synthesis in the images' background texture. The idea of producing artworks at higher resolutions would unquestionably revolutionise the audiovisual aesthetic experience. The implementation of 3D lighting would also immensely benefit from a greater computer power since those are the most demanding process performed in real-time.

As previously mentioned, another area that becomes increasingly important for future developments is the ability to include multiple participants' mode. This could be implemented in many ways. However, a virtual space available for two different participants would open a myriad of issues regarding structure, composition and arrangement. Nevertheless, the ability to include multiple participants would certainly be an interesting challenge for the creation of collaborative audiovisual aesthetic experiences.

References

- Abbado, A. (1988) *Perceptual Correspondences of Abstract Animation and Synthetic Sound*. Thesis, Massachusetts Institute of Technology.
- Adobe.com (2014) *Adobe: Creative, marketing, and document management solutions* [online]. Available at: <http://www.adobe.com> [Accessed on 19 Aug. 2014].
- Aschersleben, G. and Bertelson, P. (2003) Temporal ventriloquism: Crossmodal interaction on the time dimension. Evidence from sensorimotor synchronization. *International Journal of Psychophysiology* 50 (1–2):157–163.
- Barry, A. (2006) *Perceptual Aesthetics: Transcendent emotion, neurological image*. *Visual Communication Quarterly*. Volume 13 Summer, pp. 134-151.
- Barry, A. and Born, G. (eds.) (2014), *Interdisciplinarity: Reconfigurations of the social and natural sciences*, London: Routledge.
- Bevan, N. (2009) *Criteria for selecting methods in user-centred design*. I-USED'09 Workshop, INTERACT 2009. Uppsala, Sweden.
- Boltz, M. (2001), "Musical soundtracks as a schematic influence on the cognitive processing of filmed events", *Music Perception*, 18(4), pp. 427-454.
- Bosch, H. (1490-1510) *The Garden of Earthly Delights* [Painting]. Madrid: Museo del Prado.
- Boykov, Y. and Jolly, M. (2001) *Interactive Graph Cuts - for Optimal Boundary & Region Segmentation of Objects in N-D Images*. In the Proceedings of "International Conference on Computer Vision". Vancouver, Canada. vol.I, p.105. July.
- Bregman, A. (1990) *Auditory Scene Analysis: The Perceptual Organization of Sound*. Cambridge, Massachusetts: The MIT Press.
- Cage, J. (1957) *Experimental music: Lecture for the Convention of Music teachers National Association*.

Chang, P. (2005) Investigating Interactivity: Exploring the Role of User Power through Visual Interpretation. *In: The European Academy of Design (6th Conference), 29-31st March, Bremen, Germany*. Bremen: EAD Publishers. [Online]. Available from: http://ead.verhaag.net/fullpapers/ead06_id136_2.pdf

Cherry, C. (1953) Some Experiments on the Recognition of Speech, with One and with Two Ears". *The Journal of the Acoustical Society of America* 25 (5): 975–979.

Chion, M. (1985) *Le Son au Cinema*. Paris: Editions de l'Etoile.

Chion, M. (1994) *Audio-vision: sound on screen*, Columbia University Press, New York.

Chuang, Y., Curless, B., Salesin, D. and Szeliski, R. (2001) *A Bayesian Approach to Digital Matting*. In *Proceedings of IEEE Computer Vision and Pattern Recognition*. Vol. II, Pg. 264-271, December.

Collopy, F., Fuhrer, R. and Jameson, D. (1999) *Visual Music in a Visual Programming Language*. In *1999 IEEE Symposium on Visual Languages Tokyo, Japan*, pp. 111-118.

Cuba, Larry (1983) *Artists interfacing with technology: Basic concepts of digital creation*. *Proceeding SIGGRAPH '83 Proceedings of the 10th annual conference on Computer graphics and interactive techniques*. Pages 311-313. Detroit, Michigan, USA — July 25 - 29.

Dewey, J. (1934) *Art as Experience*. New York: The Berkley Publishing Group, Penguin.

Dinkla, S. (2002) *'The Art of Narrative'*. In *New Screen Media: Cinema/Art/Narrative*, edited by Martin Rieser, and Andrea Zapp, 105-119. London: British Film Institute.

Dix, A., Finlay, J., and Abowd, G. D. (2004) *Human-Computer Interaction*. Prentice Hall.

Dourish, P. (2001) *Where the action is: The foundations of embodied interactions*. Cambridge, MA: MIT Press.

Dykhoff, K. (2003) About the perception of sound. University College of Film, Radio, Television. Stockholm.

Edelman, G. (2007) *Second nature: Brain science and human knowledge*. New Haven: Yale University Press.

Fagundes, D. (2014) *Interactive Art Gallery: Uses of Music, Sound Design and Narrative to Explore Artworks*. Awarded as the Best Student Design Project. Presented during the 136th Audio Engineering Society Convention in Berlin. Germany, April 2014.

Fouad, H. and Hahn, J. (1998) A framework for integrating sound into Virtual Environment interfaces. Presence: Teleoperators and Virtual Environments archive. Volume 7, Issue 1. Pg 67-77.

Gibson, J. (1966) *The senses considered as perceptual systems*. Boston: Houghton Mifflin.

Gibson, J. (1979) *The ecological approach to visual perception*. Boston: Houghton Mifflin.

Goyal, P. and Diwakar, S. (2010) *Fast and Enhanced Algorithm for Exemplar Based Image Inpainting*. In Fourth Pacific-Rim Symposium on Image and Video. Singapore. IEEE. Pg 325 - 330.

Graff, H. (2015) *Undisciplining Knowledge: Interdisciplinarity in the Twentieth Century*. Baltimore: Johns Hopkins University Press.

Grierson, M. (2005) *Audiovisual Composition*. PhD Thesis. University of Kent.

Grodal, T. (2009) *Film Aesthetics and the Embodied Brain*. In Neuroaesthetics. Edited by Skov, M. and Vartanian, O. In Neuroaesthetics. New York: Baywood Publishing Company.

Hahn, J., Fouad, H., Gritz, L. and Lee, J. (1998) Integrating Sounds and Motions in Virtual Environments. Presence: Teleoperator and Virtual Environments, 7(1), 67-77.

Hales, C. (2002) 'New Paradigms \leftrightarrow New Movies.' in *New Screen Media: Cinema/Art/Narrative*, edited by Martin Rieser, and Andrea Zapp, 105-119. London: British Film Institute.

- Hayles, N. (2012). *How we think: Digital media and contemporary technogenesis*. Chicago, U.S.A.: University of Chicago Press.
- Heideman, M., Johnson, D. and Burrus, C. (1985) *Gauss and the history of the fast Fourier transform*. *Archive for History of Exact Sciences*. 34 (3): 265–277.
- Henriksen, F. (2002) *Space in Electroacoustic Music: Composition, Performance and Perception of Musical Space*. PhD Thesis. City University. Department of Music. July 2002.
- Kaneko, T., Takahei, T., Inami, M., Kawakami, N., Yanagida, Y., Maeda, T. and Tachi, S. (2001) *Detailed Shape Representation with Parallax Mapping*. In *Proceedings of ICAT 2001*, pp. 205-208.
- Kawaguchi, Y. (1987) *Ecology II: Float II* [Computer animation]. *Ars Electronica* [Online]. Available at: <http://archive.aec.at/showmode/prix/?id=20746#20746>
- Kay, A. (1984) *Computer Software*. *Scientific American*, 251, 52– 59.
- Lakoff, G., and M. Johnson (1983) *Metaphors We Live By*. Chicago: University of Chicago Press.
- Lakoff, G. (2013a) *What Studying the Brain Tells Us About Arts Education* [Online]. Available at: <http://www.youtube.com/watch?v=fpla16Bynzg>
- Lakoff, G. (2013b) *Embodied Cognition and Language* [Online]. Available at: <http://www.youtube.com/watch?v=XWYaoAoijdQ>
- Latour, B. (1999) *Pandora's Hope: Essays on the reality of sciences studies*. Cambridge, MA: Harvard University Press.
- Latour, B. (2005) *Reassembling the Social: An Introduction to Actor-Network-Theory*. UK: Oxford University Press.
- Laurel, B. (1993) *Computers as Theatre*. New York: Wesley Longman, Inc.
- Lipscomb, S. (2005) *The perception of audio-visual composites- Accent structure alignment of simple stimuli*. *Selected Reports in Ethnomusicology* 12: pp. 37-67.

McLaren, Norman (1976) *Animated Sound on Film in Experimental Animation: An Illustrated Anthology*. Robert Russett and Cecile Starr. New York: Van Nostrand Co.

Manovich, L. (2001) *The Language of New Media*. Cambridge, Mass: MIT Press.

Manovich, L. (2002) 'Spatial Computerisation and Film Language.' in *New Screen Media: Cinema/Art/Narrative*, edited by Martin Rieser, and Andrea Zapp, 64 - 76. London: British Film Institute.

Marclay, C. (2015) *Surround Sounds* [Art Instalation]. Presented on the 27th Jan 2015 at White Cube Art Gallery. London, United Kingdom.

McLuhan, M. (1959) *Electronic Revolution: Revolutionary Effects of New Media*. In *Understanding Me: Lectures and Interviews*. Edited by Stephanie McLuhan and David Staines (2003). New York: MIT Press.

McLuhan, M. (1960) *Popular / Mass Culture: American Perspectives*. In *Understanding Me: Lectures and Interviews*. Edited by Stephanie McLuhan and David Staines (2003). New York: MIT Press.

McLuhan, M. (1964a) *Understanding Media: The Extensions of Man*. London: Routledge and Kengan Paul Limited.

McLuhan, M. (1964b) *Cybernetics and Human Culture*. In *Understanding Me: Lectures and Interviews*. Edited by Stephanie McLuhan & David Staines. Toronto: McClelland & Stewart.

Meikle, G. (2002) *Future Active: Media Activism and the Internet*. London: Routledge and Sydney: Pluto Press.

Moran, J. (2010) *Interdisciplinarity: The New Critical Idiom*. London: Routledge.

Mortensen, E. and Barrett, W (1998) *Interactive segmentation with intelligent scissors*. In *Graphical models and image processing*. Vol. 60. Issue 5. Pg. 349-384. Academic Press.

Murray, J. (1997) *Hamlet on the Holodeck: the future of narrative in cyberspace*. Cambridge, Massachusetts: MIT Press.

Murray, J. (2011) *Inventing the medium: Principles of interaction design as a cultural practice*. Cambridge, Massachusetts: MIT Press.

Murray, S. (1995) *Engaging Characters: Fiction, Emotion, and the Cinema*. Oxford: Clarendon Press.

Nixon, M. and Bizzocchi, J. (2013) *Press X for Meaning: Interaction Leads to Identification in Heavy Rain*. School of Interactive Arts & Technology, Simon Fraser University.

Noble, J. (2012) *Programming Interactivity*. USA: O'Reilly Media.

Noë, A. (2006) *Action in Perception*. Cambridge: MIT Press.

Noë, A. (2010) *Out of Our Heads: Why You Are Not Your Brain, and Other Lessons from the Biology of Consciousness*. New York: Hill and Wang.

Norman, D. (1988). *The Design of Everyday Things*. Basic Books: New York.

Papert, S. (1980) *Mindstorms: Children, Computers, and Powerful Ideas*. New York: Basic Books.

Parhi, M., Lao, Y. and Keshab, K. (2014) *Canonic Real-Valued FFT Structures*. Department of Electrical and Computer Engineering. University of Minnesota.

Processing.org (2014) *Processing 2: programming language, development environment, and online community* [Online]. Available at: <http://processing.org> (Accessed on 25 Aug. 2014).

Puckette, M. (1996) *Pure Data*. Proceedings, International Computer Music Conference. San Francisco: International Computer Music Association, pp. 224-227.

Ramachandran, V.S. (2011) *The Tell-Tale Brain: Unlocking the mystery of human nature*. London: Heinemann.

Rieser, M. and Zapp, A. (2002) *New Screen Media: Cinema/Art/Narrative*. London: British Film Institute.

Rizzolatti G, Luppino G. (2001) The cortical motor system. *Annu. Rev. Neurosci.* 27:169–92.

Rizzolatti G, Craighero L. (2004) The mirror-neuron system. *Annu. Rev. Neurosci.* 27:169–92.

Robart, R. and Rosenblum, L. (2005) Hearing space: Identifying rooms by reflected sound. In H. Heft and K. L. Marsh, eds., *Studies in Perception and Action XIII*. Hillsdale, NJ, USA: Lawrence Erlbaum.

Rother, C., Kolmogorov, V. and Blake, A. (2004) *GrabCut: interactive foreground extraction using iterated graph cuts*. In *ACM Transactions on Graphics (SIGGRAPH)*. Vol. 23.

Russett, R. and Starr, C. (1988) *Experimental Animation: Origins of a New Art*. Revised.

Russolo, L. (1913) *The Art of Noises*: (Futurist Manifesto, 1913). In 'The art of noises': A great bear pamphlet Series, 1967. New York: Something Else Press.

Salter, C. (2010) *Entangled: Technology and the transformation of performance*. Cambridge, Mass: MIT Press.

Schaeffer, P. (1966) *Traite' des objets musicaux*. Paris: E' ditions du Seuil.

Sekuler, R. and Blake, R. (1985) *Perception*. New York: McGraw-Hill.

Shapiro, L. and Stockman, G. (2001) *Computer Vision*. New Jersey: Prentice-Hall.

Shimamura, A. (2012) Aesthetic Science: Connecting Minds, Brains and Experiences. In A. P. Shimamura & S. E. Palmer (Eds.), (pp. 3-28). Oxford: Oxford University Press.

Sorensen, V. (2014) *Conceptions of Space-Time in Illuminations, an Animated Installation*. 26th Society for Animation Studies (SAS) Conference, Toronto, Canada, June 17.

Spaulding, E. and Faste, H. (2013) *Design-driven narrative: using stories to prototype and build immersive design worlds*. In: *Proceedings of the SIGCHI*

Conference on Human Factors in Computing Systems (CHI 2013), pp. 2843–2852. ACM, New York.

Strutt, J. W. (1907) *On our perception of sound direction*, Philos. Mag., vol. 13, pp. 214-232.

Suchman, L. (2006) *Human-Machine Reconfigurations: Plans and Situated Actions (Learning in Doing: Social, Cognitive and Computational Perspectives)*. Cambridge: Cambridge University Press.

Sulonen, P. (2011) *Mastering Wide-Angle*. In Petteri's Pontifications [Online]. Available at: http://www.prime-junta.net/pont/How_to/m_Mastering_Wide-Angle/m_Mastering_Wide-Angle.html (Accessed on 31 Aug. 2014).

Vertov, D. (1929) *Man with a Movie Camera* [Film]. USSR: M. Kaufman.

Wang, W. and Liu, H. (2009) *Study of System Model of Image Inpainting Combining Subjective Estimation and Impersonal Estimation*. Proceedings of the International Multi Conference of Engineers and Computer Scientists. Hong Kong, March, 2009, pp869-873.

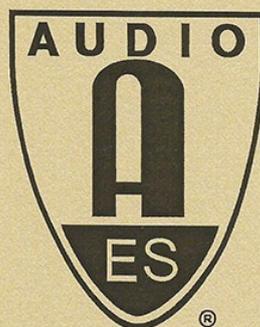
Wei, L., Lefebvre, S., Kwatra, V. and Turk, G. (2009) *State of the Art in Example-based Texture Synthesis*. In Eurographics 2009. State of the Art Report. EG-STAR.

Wilcox, F. (1956) *Forbidden Planet* [Film]. MGM.

Willemen, P. (2002) *'Reflections of Digital Imagery: Of Mice and Men.'* in *New Screen Media: Cinema/Art/Narrative*, edited by Martin Rieser, and Andrea Zapp, 14 - 26. London: British Film Institute.

Whitney, J. (1980) *Digital Harmony: On the Complementarity of Music and Visual Art*. Peterborough: McGraw-Hill.

Appendix A [LiveRoom]



Audio Engineering Society

AES 134th Convention
Rome, Italy
May 4-7, 2013

Student Design Competition

Diego Fagundes

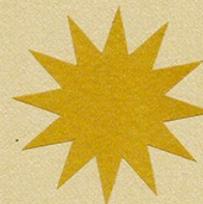
SAE Institute London

UNDERGRADUATE CATEGORY

GOLD AWARD

*“LiveRoom - An Interactive Virtual Music Listening
Environment”*

John Krivit
Education Committee Chair



LiveRoom*

An Interactive Virtual Music Listening Environment



Figure 1. The LiveRoom (GEM video display)*

ABSTRACT

This paper presents the LiveRoom*, a virtual music listening environment that engages users in an interactive experience with recorded music and moving images. The project aims to virtually reproduce an authentic musical event integrating video and audio processing.

I. INTRODUCTION

This project consists of a music playback tool that offers an entertaining experience, and at the same time an educational contribution to the music industry. The LiveRoom* places the users in the centre of a virtual room, enabling them to act as producers by navigating inside it and immediately changing the audible response.

The goal is to convey acoustical messages from the sound sources to the receiver, as it would happen in a real-world scenario.

For the purpose of this project, audio and video footage of three songs performed by Citizen were captured. In order to achieve accurate results, several stereo microphone techniques and close microphone techniques were employed during the recording process, and, ultimately, combined in the mixing stage.

II. CONCEPT

The LiveRoom* was intended to provide an alternative music listening experience. As the listener navigates through the immersive listening environment, the distinct sound sources are subjected to several changes in level, balance, reverberation, panning and equalisation. Accordingly, artificially recreating the listener's ability to identify the localisation and origin of a produced sound.

This project was designed considering vital psychoacoustic factors, such as the audio-visual contract and the auditory scene analysis. Although the project required the programming of a complex processing system in order to accomplish these particular results, the user's control is very intuitive. The listener interacts by moving the cursor across the user interface, undertaking a perceptual and cognitive experience.

III. THE LIVEROOM* DESIGN

The listener position is represented by a 3D model of a dummy head, which can be rotated and moved inside the virtual room in real-time, generating an audio-visual relation between the musicians' performance and the audible response.



Figure 2. Focusing on the keyboards

For instance, by placing the dummy head closer to one of the musicians, the user can visually and sonically focus on a specific musical instrument, changing its level, reverberation, panning and frequency response. Simultaneously, all the other instruments have their audio response also affected, interpreting the listener's perspective in relation to the sound sources.

The 3D model dummy head position is limited to a rotational range of 45° between both extremes of the room, so as to keep the listeners constantly facing the sound sources.

Although all instruments performed during the recording are permanently present in the mix, their individual sonic response and the overall level is constantly changing as the head is rotated and allocated to different positions in the virtual space. The audio output can be reproduced through either headphones or loudspeakers.

A. Programming in Pure Data

The programming of the application was performed in Pure Data language using the Pd-Extended software, where all the individual audio stems, videos, and 3D objects were introduced, and the audio processing and interactive system were implemented.

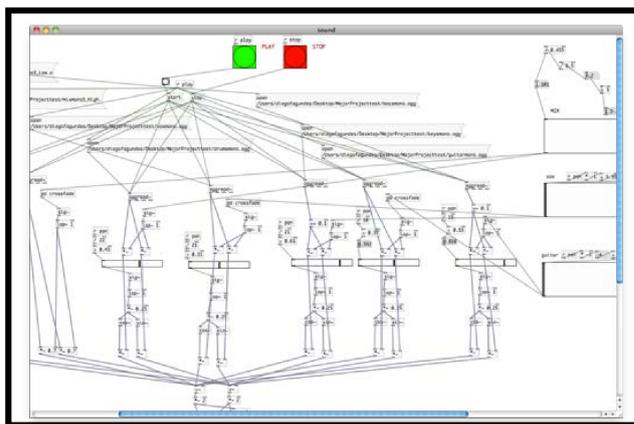


Figure 3. Pure Data patching (audio section)

The programming process was inspired by several projects previously proposed and, subsequently, various theories, ideas, and concepts were combined, modified and adapted in order to meet the application's needs.

The first step in the programming process was to create an audio player and a video player independently, and synchronizing them. This was achieved with the use of the objects [readsf~] and [pix_film].

All the audio and video files were triggered simultaneously to certify a perfect synchronization. The visual content was displayed and manipulated through the use of the Graphics Environment for Multimedia (GEM) window, responsible for providing graphics functionality to Pure Data (PD Community Site, 2013).

B. Audio Processing

The LiveRoom* contains a series of interconnected effects and processing that emulate fundamental elements associated to the human hearing

process.

As an example, a number of crossfades and Low/High pass filters recreate the interaural intensity difference by reconstructing in real-time a variable difference in loudness between the two ears, based on the movements controlled by the users, rotating the three-dimensional head object. Reverberation and other time based effects, are employed to emulate the Interaural time difference, which can be described as the difference in time that sound takes to reach both ears (Farnell, 2008).

C. User Interface

The user interface was designed to function as the platform for interaction. The [grid] object displayed in the centre allows the user to control all the audio and visual processing. Alternatively, it can be remotely operated with the use of a touch sensitive device running an OSC application.

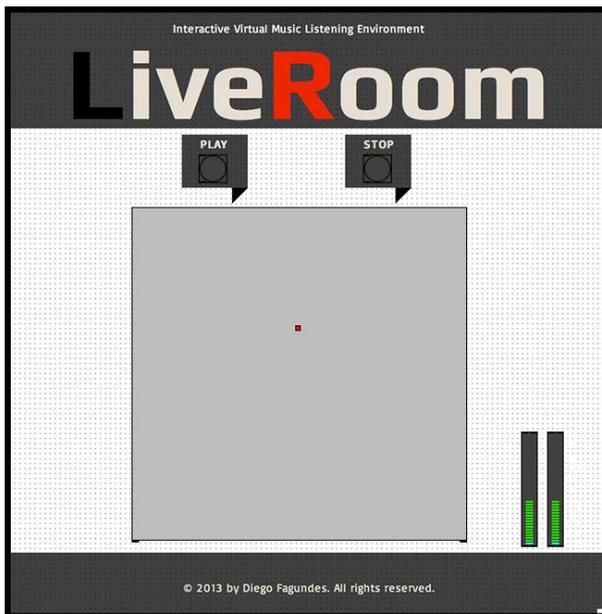
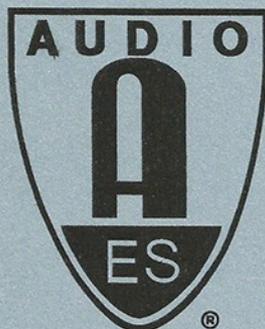


Figure 4. User Interface, XY [grid] object.

Bibliography

- A. Farnell, "Designing Sound". London: Applied Scientific Press, 2008.
- A. Copland, "What to listen for in music." London: Signet Classics, 2002.
- B. Owsinsky, "The recording engineer's handbook: second edition." USA: Artistpro Publishing, 2006.
- D. Levitin, "This is your brain on music." London: Atlantic Books, 2006.
- E. Huhtamo, "From Cybernation to Interaction: A Contribution to an Archaeology of Interactivity". The Digital Dialectic: New Essays on New Media. Cambridge, MA: MIT Press, 1999.
- F. Rumsey, "Spatial Audio." Oxford: Focal Press, 2001.
- H. Peng, C. Chou, and C. Chang, "From Virtual Environments to Physical Environments: Exploring Interactivity in Ubiquitous-learning Systems' Educational Technology & Society, 2008.
- J. Schnupp, I. Nelken and A. King, "Auditory neuroscience: Making sense of sound". Cambridge, Massachusetts: The MIT Press, 2011.
- J. Douglas, and A. Hargadon, "The Pleasure of Immersion and Interaction: Schemas, Scripts, and the Fifth Business", Cambridge MIT Press, 2004.
- L. Yuping and L. Shrum, "What is interactivity and is it always such a good thing? Implications of Definition, Person, and Situation for the Influence of Interactivity on Advertising Effectiveness." Journal of Advertising. 31, 4. Abi/Inform Global, 2002.
- M. Wilde, "Audio Programming for Interactive Games." Oxford: Focal Press, 2004.
- O. Grau, "Virtual Art: From Illusion to Immersion." Cambridge, MA: MIT Press, 2004.
- PD Community Site "List Of Gem Objects" [Online], 2013.
- PD Community Site, "Graphics Environment for Multimedia" Project Description [Online], 2013.
- S. McMillan, "Exploring Models of Interactivity from Multiple Research Traditions: Users, Documents, And Systems." University of Tennessee, 2000.
- S. Brown, "Smoke and Mirrors" [Interactive Artwork], no date.

Appendix B [12 Tones]



Audio Engineering Society

AES 137 Convention
Los Angeles
October 9th-12th, 2014

Student Design Competition

Diego Fagundes

Goldsmiths, University of London

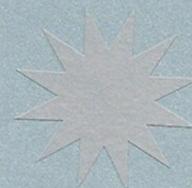
Graduate Category

SILVER AWARD

“12 Tones”



John Krivit
Education Committee Chair



12 Tones*

Real-time Sound Visualisation for Live Performance



Figure 1. Live performance with 12 Tones* system at Kingston University, Kingston Hill Auditorium in London, on the 19th May 2014 (Fagundes, 2014).

ABSTRACT

The idea of combining image and sound in live performances is not new. Nevertheless, current advances in digital technologies have made this concept more tangible. This paper presents 12 Tones*, an application that analyses particular proprieties of audio signals and apply it to the generation and direct control of visual material (Figure 1). This project associates real-time processing and interactivity into a single tool, designed to create multimedia narratives in live performance scenarios.

I. INTRODUCTION

There is an on going interest in the creation of instruments to play graphics in the same way that musicians play sounds (Collopy et al., 1999). The notion of linking sound and image can be tracked back to Greek philosophy (Moritz, 1997), like Aristotle and Pythagoras, who have contemplated the association between musical scales and the hue spectrum of the rainbow (Ciufu, 2002). In *De Sensu* Aristotle observes this relationship:

“Colours may mutually relate like music concords for their pleasantest arrangement; like those concords mutually proportionate” (as cited in Wilfred, 1947).

In 1725 the French mathematician Louis Bertrand Castel crafted his *Clavecin Pour Les Yeux*, an ocular harpsichord inspired by Newton’s investigation on colour theory. Subsequently, there have been various experimentations developed by music composers, painters and filmmakers (Collopy et al., 1999).

II. CONCEPT

Sound visualisation applications can be grouped into two distinct categories. The first comprises image sequencers, which provide playback, transitions and effects to moving images previously produced. The second category is comprised by image generators. These applications permit the creation of graphics, often generated to complement musical material (Collopy et al., 1999).

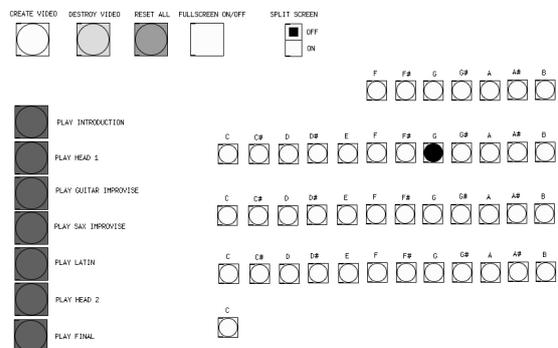


Figure 2. 12 Tones* Graphic User Interface in Pd-extended.

This project integrates both approaches, enabling the performer to generate graphics in real-time, as well as, to control the movement and behaviour of images previously created. 12 Tones* generates visual material



Figure 5. 12 Tones* opening scene, where the lights are controlled by the guitar's amplitude signal.

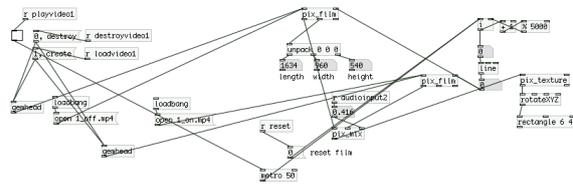
The second part of the visual system is created using the [pix_film] object. The [pix_film] object loads videos and applies them to the Graphics Environment for Multimedia [GEM] as textures (Figure 5). Following, the transition between the different images is controlled by the audio signal's amplitude using the [pix_mix] object, which mixes two images based on mixing factors (Figure 6). The mixing factors (A and B) are controlled by the values received from the [env~] object, resulting in the output: $out=in1*A+in2*B$.

IV. CONCLUSION

12 Tones* is a real-time sound visualisation tool that can be applied to any musical instrument. In this project, we explore the relationship between sound and colour. However, the particular aesthetic choices are left for the system designer and the performer, as the primary purpose of this project is to propose a tool for audio and visual association.

The system's architecture has been described to demonstrate its potential for generating and controlling dynamic visuals, consisting of a powerful tool in the hands of musicians and artists. The usefulness of this framework could undoubtedly be increased with the development of a method to include multiple instruments to control the graphic material. This is a fundamental area for further investigation. Moreover, future work will include the implementation of 3D models, as well as, a deeper level of interaction between performers and the system.

The notion of colour music is not novel; therefore, it can be significantly deliberated with the examination of its history. The new technology and tools currently



Correia, N. (2012) AVOL: Towards an Integrated Audio-Visual Expression. *Journal of Visual Art Practice*. Volume 10, Issue 3. Pg 201-214.

PD Community Site (2014) Graphics Environment for Multimedia. Project Description [Online]. Available at: <http://gem.iem.at/>

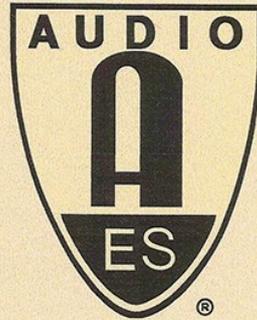
PD Community Site (2014) List Of Gem Objects [Online]. Available at: http://gem.iem.at/documentation/manual/manual/list-of-gem-objects/?searchterm=pix_draw

Pure Data Forum (2014) Pure Data forum~ [Online]. Available at: <http://puredata.hurlleur.com/recherche.html>

Vaughan, T. (2008) *Multimedia: Making it Work*. 7th Ed. London: McGraw-Hill.

Wilfred, T. (1947) Light and the Artist. *Journal of Aesthetics and Art Criticism*, June.

Appendix C [Interactive Art Gallery]



Audio Engineering Society

AES 136 Convention
Berlin, Germany
April 26 - 29, 2014

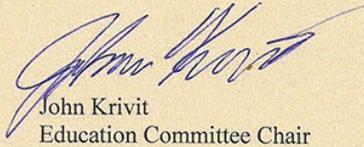
Student Design Competition

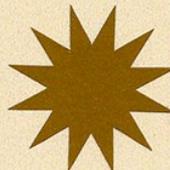
Diego Fagundes

Goldsmiths, University of London

**GRADUATE
GOLD AWARD**

“INTERACTIVE ART GALLERY”


John Krivit
Education Committee Chair



Interactive Art Gallery*

Uses of Music, Sound and Narrative to Explore Artworks

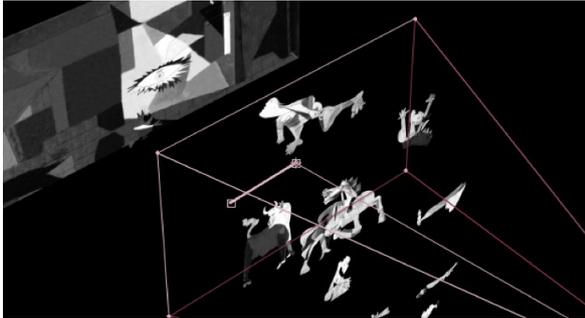


Figure 1. An exploration of Pablo Picasso's 'Guernica' (IAG* video display)

ABSTRACT

This paper presents the Interactive Art Gallery*, a multimedia interactive tool, with an entertaining and educational purpose, intended to offer an alternative point of view of an artwork, enriched with music, sound textures, narrative and three-dimensional movements.

I. INTRODUCTION

The main intention of this project is to implement sound and interactivity to augment artworks, proposing an educational tool that sonically and visually illuminates various events and aspects related to its artistic conception. Moreover, the aural-visual relation and the sense of three-dimensional immersion will provide an opportunity for the user to interact and observe a piece of art from a distinctive perspective.

The project envisages exploring and incorporating diverse visual art forms, including painting, photography, sculpture, graphic design, film, documentary and dance.

II. CONCEPT

Interactive Art Gallery* project explores the use of music, sound design, sonic textures and voices in the form of narratives to reveal facets that would usually stay unseen by casual viewers. Furthermore, IAG* visually fragments the work of art, so as to present different analyses designed to help in the recognition of features that most expressively constitute it. All information is displayed as a multimedia playlist seamlessly blended to each other. It can be interactively accessed through the graphical user interface in an intuitive way, as the project is targeted to complete novices, with no previous familiarity.

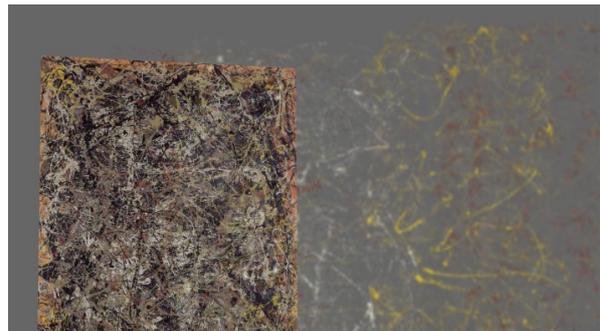


Figure 2. An exploration of Jackson Pollock's No. 5 [1948] (IAG* video display)

III. THE INTERACTIVE ART GALLERY* DESIGN

This section of the paper will uncover all the stages of development and programming of the IAG* as well as the project's management aspects.

A. Still Images

The development of the project's visual content begins with the manipulation of still images. The intent is to create virtual camera movements with 3D depth, manipulating only still images as the basic source. Firstly, the chosen symbols or characters are extracted from the background of the artwork with the use of mask selections, turning them into new individual layers. The next step is to fill the gaps behind the layers, preventing the existence of dead spaces in the background image.

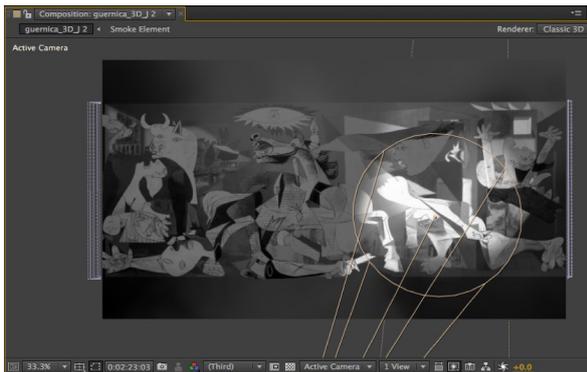


Figure 3. Video editing session for Picasso's 'Guernica' (IAG exploration)*

B. The 3D Effect

Firstly, the composition of layers is arranged into 'z-space', in order to create the three-dimensional effect. Following, motion and animation are applied to the composition and a 35mm camera simulation is created permitting the design of a 'flying' movement across the different layers, creating a sense of immersion and adding a theatrical quality to the animation.

C. Recording and Mixing Process



Figure 4. Voice-Over Recording feat. Pedro Polônio

The voice-over recording process focuses on the creation of a sonic conception, proposed to not only outline information, but also, to establish a relationship between the picture's movements and the narrative. The final mix is performed in Pro Tools where the audio and the footage are processed and synchronized.



Figure 5. IAG final mix Pro Tools session*

- Huhtamo, E. (1999) *From Cybernation to Interaction: A Contribution to an Archaeology of Interactivity*. The Digital Dialectic: New Essays on New Media. Cambridge, MA: MIT Press [Online]. Available at: <http://classes.dma.ucla.edu/Fall06/10/CybernationToInteraction.pdf>
- Levitin, D. (2006) *This is your brain on music*. London: Atlantic Books.
- Manovich, L. (2006) *What is new media? The New Media Theory Reader*. London: Open University Press.
- McMillan, S. (2000) 'Exploring Models of Interactivity from Multiple Research Traditions: Users, Documents, And Systems.' University of Tennessee [Online]. Available at: <http://online.sfsu.edu/ottasca/beca300/secure/mcmillan.pdf>
- Murray, J. (1999) 'Interactive Design: A Profession in Search of Professional Education'. The Chronicle of Higher Education [Online]. Available at: <http://faculty.winthrop.edu/kosterj/WRIT501/readings/MurrayInteractiveDesign1999.htm>
- Negroponte, N. (1995) *Being Digital*. London: Hodder and Stoughton.
- Peng, H., Chou, C., and Chang, C. (2008) 'From Virtual Environments to Physical Environments: Exploring Interactivity in Ubiquitous-learning Systems' Educational Technology & Society [Online]. Available at: http://www.ifets.info/journals/11_2/6.pdf
- PD Community Site (2013) Graphics Environment for Multimedia. Project Description [Online]. Available at: <http://gem.iem.at/>
- PD Community Site (2013) List Of Gem Objects [Online]. Available at: http://gem.iem.at/documentation/manual/manual/list-of-gem-objects/?searchterm=pix_draw
- Pure Data Forum (2013) Pure Data forum~ [Online]. Available at: <http://puredata.hurlleur.com/recherche.html>
- Vaughan, T. (2008) *Multimedia: Making it Work*. 7th Ed. London: McGraw-Hill.
- Yuping, L. and Shrum, L. (2002) 'What is Interactivity and is it Always Such a Good Thing'. Journal of Advertising. Winter 2002 [Online]. Available at: http://www.yupingliu.com/files/papers/liu_shrum_interactivity.pdf

Appendix D [User Study Questionnaire]

AVC Image Segmentation and Inpainting

An automated system combining image segmentation and image inpainting

1. General Information

1.1. Age

1.2. Gender

Male

Female

1.3. What is your educational background?

1.4. What is your current occupation?

1.5 Do you have experience as visual artist or designer?

Yes, as visual artist

Yes, as designer

Yes, as both

1.6. (Related to the previous question) For how many years?

2. Have you used any other image segmentation application before?

Yes.

No.

2.1 (If you answered yes) which one?

3. Have you used any other image inpainting (or texture synthesis) application before?

Yes.

No.

3.1 (If you answered yes) which one?

4. With regards to foreground/background separation, how satisfied are you with the image segmentation results?

1 2 3 4 5

Not at all satisfied Extremely satisfied

5. How satisfied are you with the resulted background image after the automated inpainting applied?

1 2 3 4 5

Not at all satisfied Extremely satisfied

6. In terms of composition, is the background / middle ground / foreground arrangement correct?

1 2 3 4 5
Totally inaccurate Extremely Accurate

7. Do you consider an automated application (with no adjustable parameters) suitable for the task of segmenting still images?

1 2 3 4 5
Not suitable Very suitable

8. How satisfied are you with an application that integrates image segmentation and image inpainting in a unique automated system?

1 2 3 4 5
Not at all satisfied Extremely satisfied

9. Would it be important for you to have control over the number of layers resulting from the image segmentation process?

1 2 3 4 5
Not at all important Extremely important

10. Are the functionalities of the application clear enough?

1 2 3 4 5
Not at all clear Extremely clear

11. Would you use AVC Image Segmentation as a new tool for your own art practice?

Yes.

No.

12. (Related to the previous question) why?

13. Does the application behave as you expect, when you operate it?

1 2 3 4 5

It does not behave at all as I expected

<input type="checkbox"/>				
--------------------------	--------------------------	--------------------------	--------------------------	--------------------------

It behaves exactly as I expected

14. Do you feel that you have control over the image segmentation process?

1 2 3 4 5

I feel I have no control at all

<input type="checkbox"/>				
--------------------------	--------------------------	--------------------------	--------------------------	--------------------------

I feel I have total control over it

15. Do you consider that you are creating your own artwork while using the application?

1 2 3 4 5

No, I don't consider it my own artwork

<input type="checkbox"/>				
--------------------------	--------------------------	--------------------------	--------------------------	--------------------------

Yes, I feel that I am creating my own work

16. Do the images generated by the application inspire your creativity?

Yes.

No.

17. (Related to the previous question) why?

18. Do you think the AVC Image Segmentation would be suitable as a plugin for graphics editor software (eg. Adobe Photoshop)?

1 2 3 4 5

Not suitable at all Yes, extremely suitable

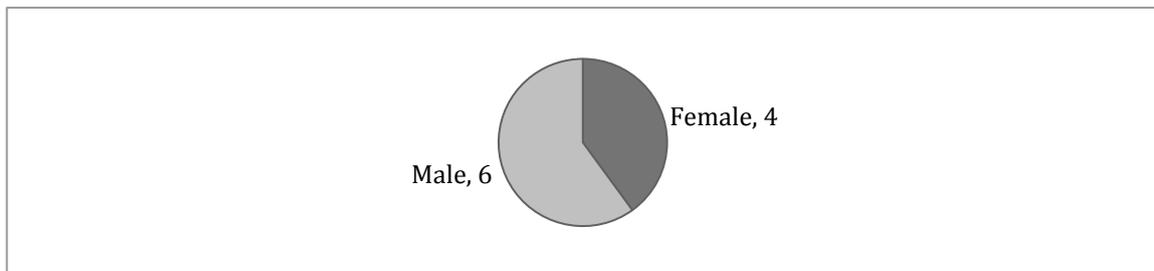
19. Do you have any thoughts on how to improve this application?

Responses to the User Study Questionnaire

- **Age**

P1. 33 P2. 35 P3. 30 P4. 37 P5. 33 P6. 28 P7. 35 P8. 34 P9. 41 P10. 52

- **Gender**



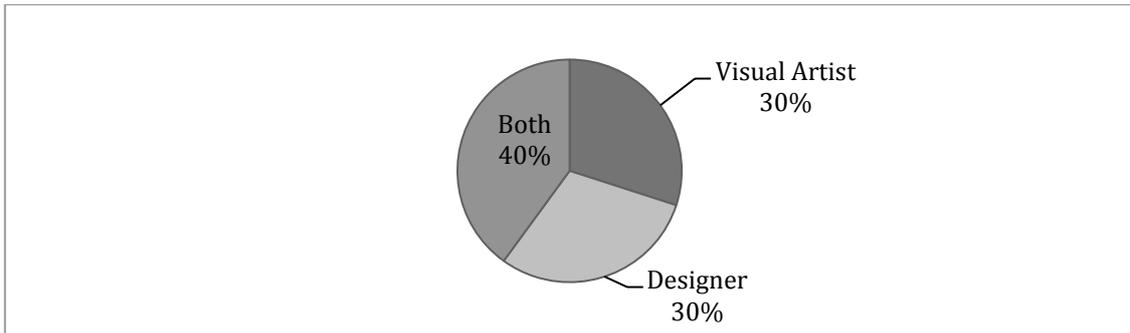
- **What is your educational background?**

- P1. Event Management & Graphic Design (MA)
- P2. Media & Advertising Degree (BA)
- P3. Music Production for Media Degree (BA)
- P4. Computational Arts (MA)
- P5. Sound Engineering (BA)
- P6. Event Management (Course)
- P7. Web Design (BA)
- P8. Digital Cinema (Higher Diploma)
- P9. Architecture (BA)
- P10. Cinema (Course)

- **What is your current occupation?**

- P1. Music Agent and Freelance Graphic Designer
- P2. Music and Art Educator
- P3. Music Performer and Composer for Media
- P4. Freelance Visual Artist
- P5. PhD Candidate in Digital Arts & Humanities
- P6. Events Coordinator
- P7. Web Designer
- P8. Music Composer
- P9. Visual Artist
- P10. Digital Marketing Manager

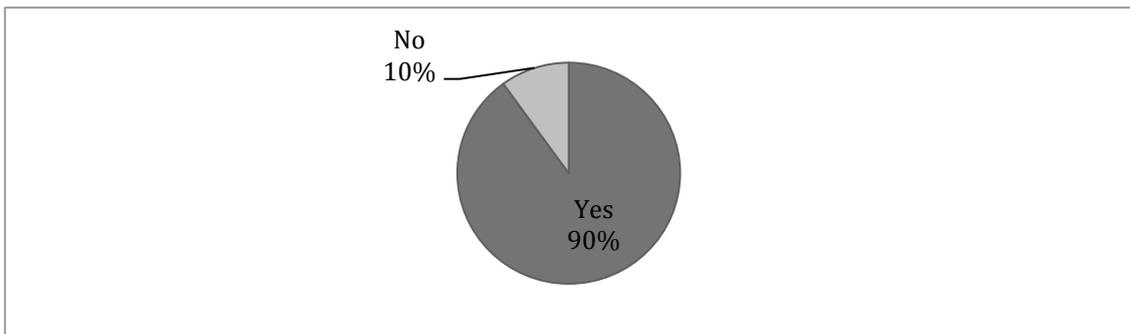
- **Do you have experience as visual artist or designer?**



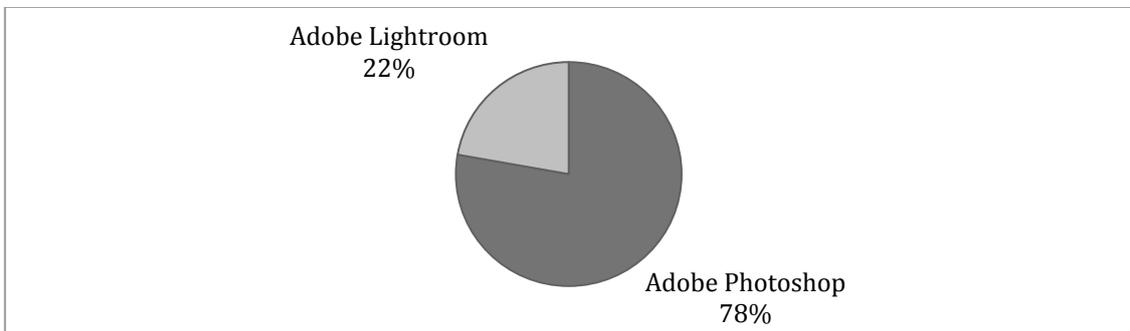
- **Have you used any other image segmentation or inpainting application before?**

Yes – P1, P2, P3, P4, P5, P6, P7, P9, P10

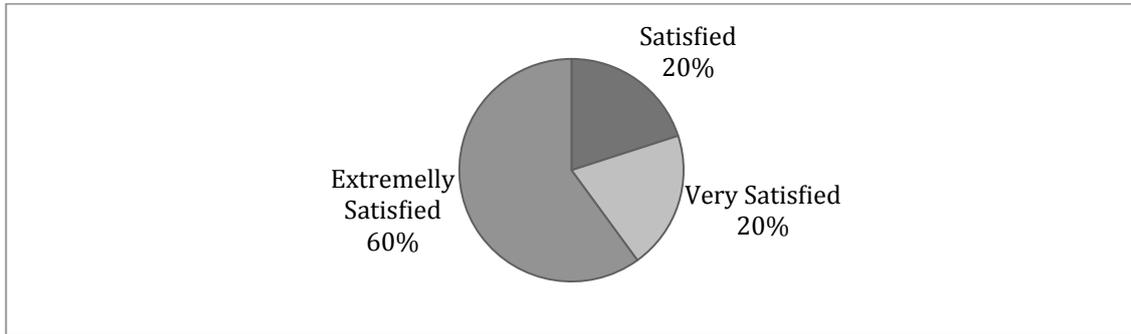
No – P8



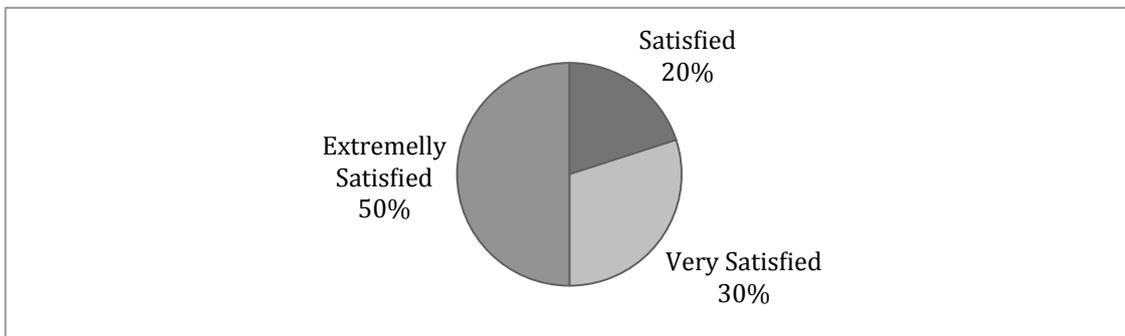
- **(If you answered yes) which one?**



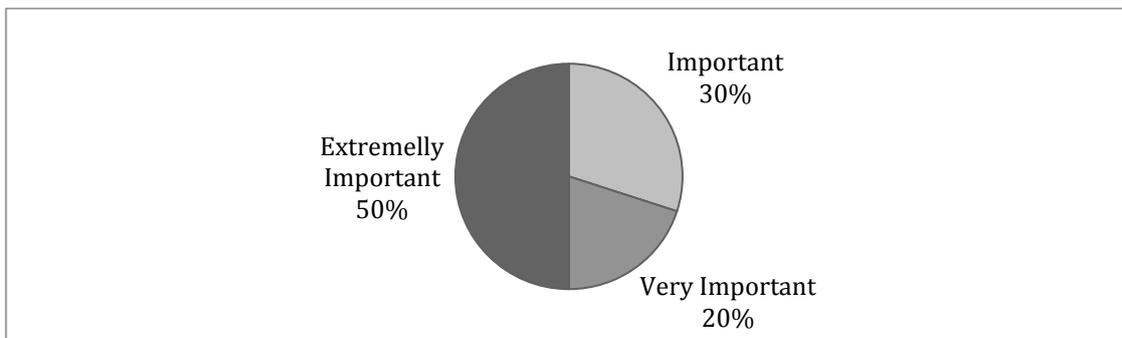
- **With regards to foreground/background separation, how satisfied are you with the image segmentation results?**



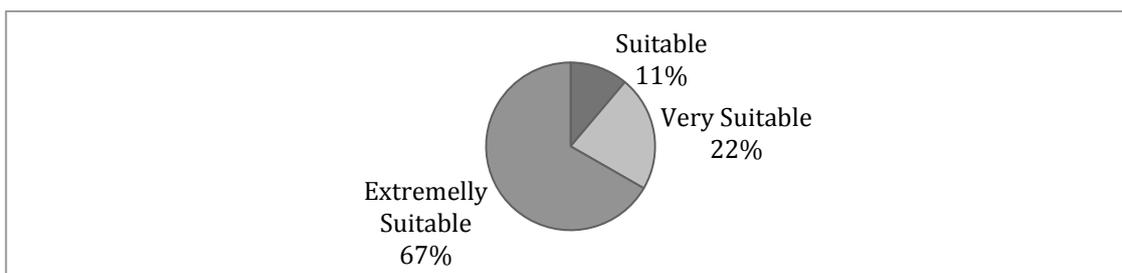
- **How satisfied are you with the resulted background image after the automated inpainting applied?**



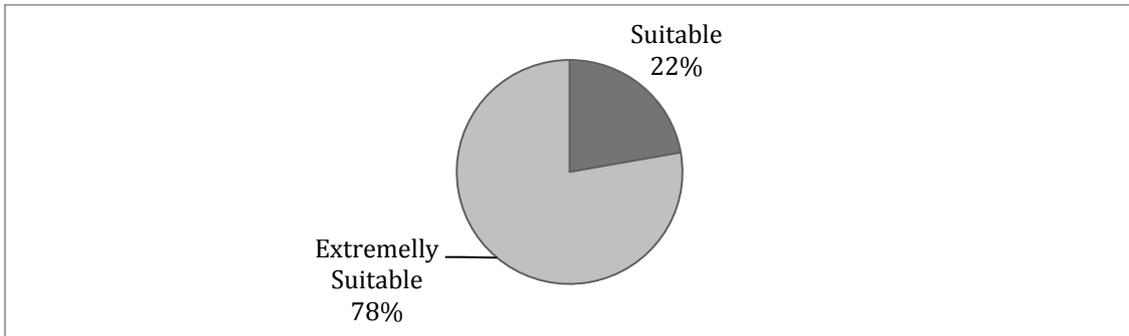
- **In terms of composition, is the background / middle ground / foreground arrangement correct?**



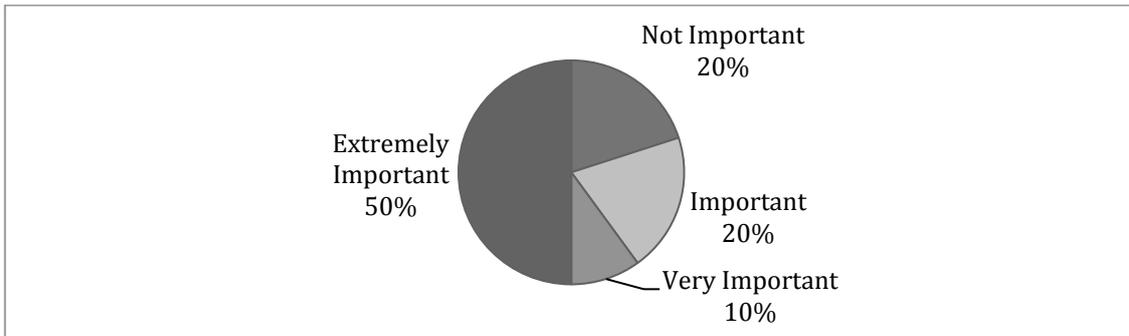
- **Do you consider an automated application (with no adjustable parameters) suitable for the task of segmenting still images?**



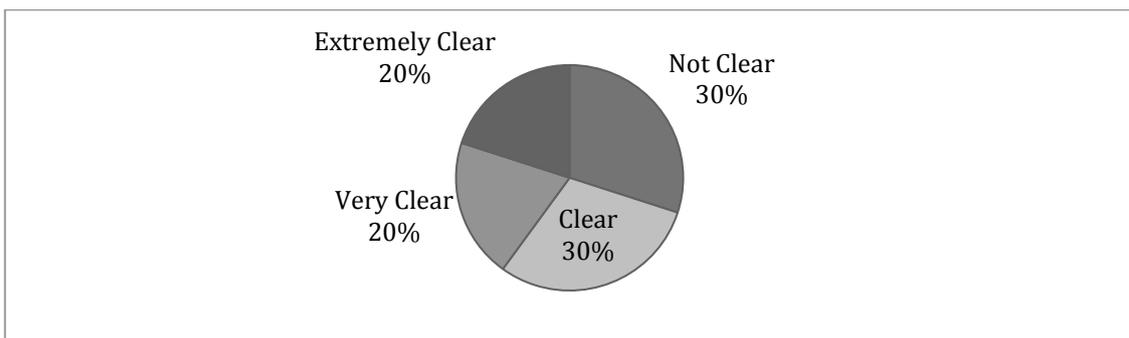
- **How satisfied are you with an application that integrates image segmentation and image inpainting in a unique automated system?**



- **Would it be important for you to have control over the number of layers resulting from the image segmentation process?**



- **Are the functionalities of the application clear enough?**

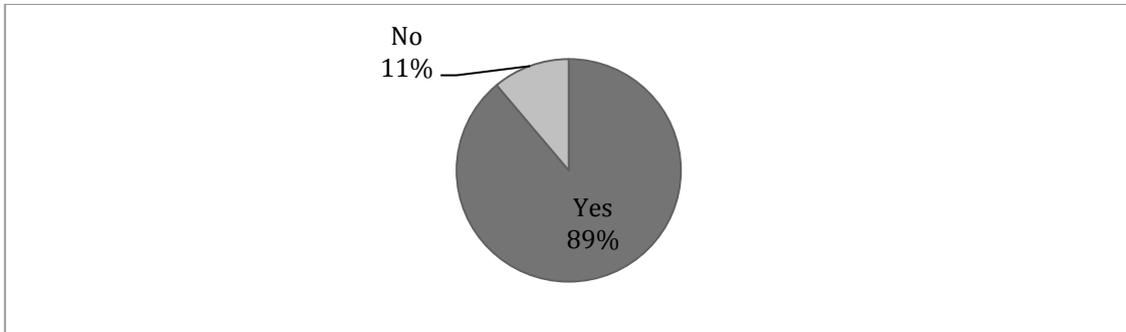


- **Would you use AVC Image Segmentation as a new tool for your own art practice?**

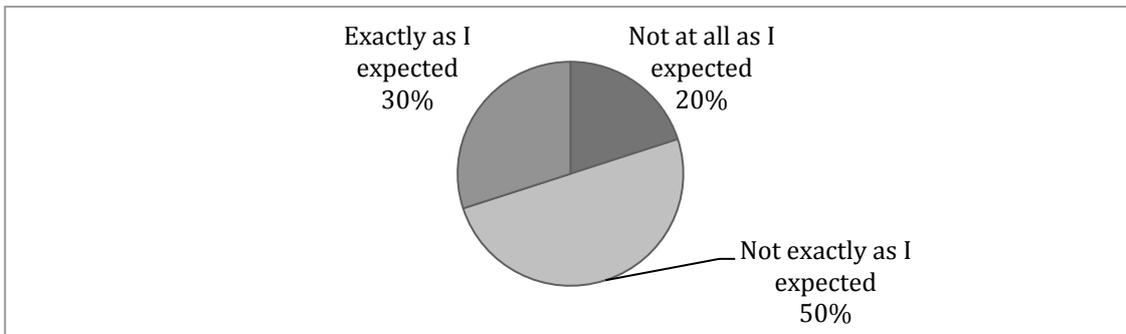
Yes - P1, P3, P4, P5, P6, P9, P8, P10

No - P9

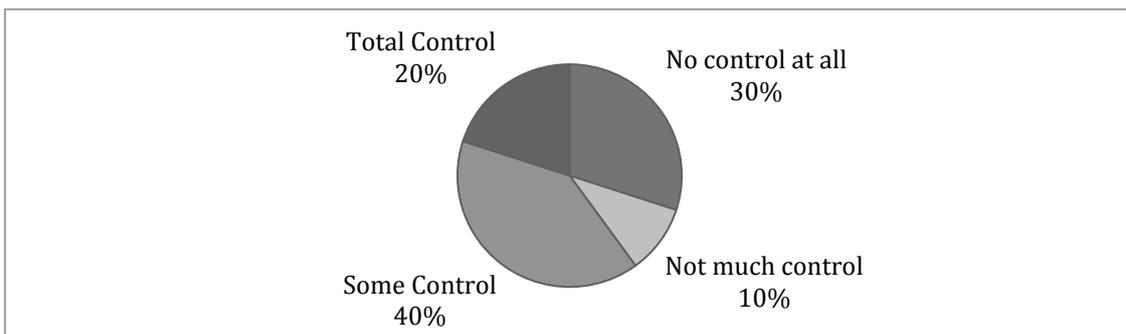
Did not respond – P2



- **Does the application behave as you expect, when you operate it?**



- **Do you feel that you have control over the image segmentation process?**

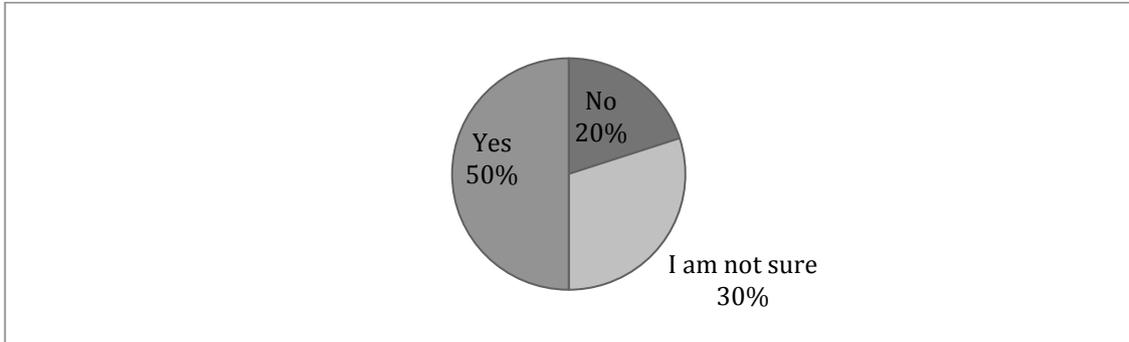


- **Do you consider that you are creating your own artwork while using the application?**

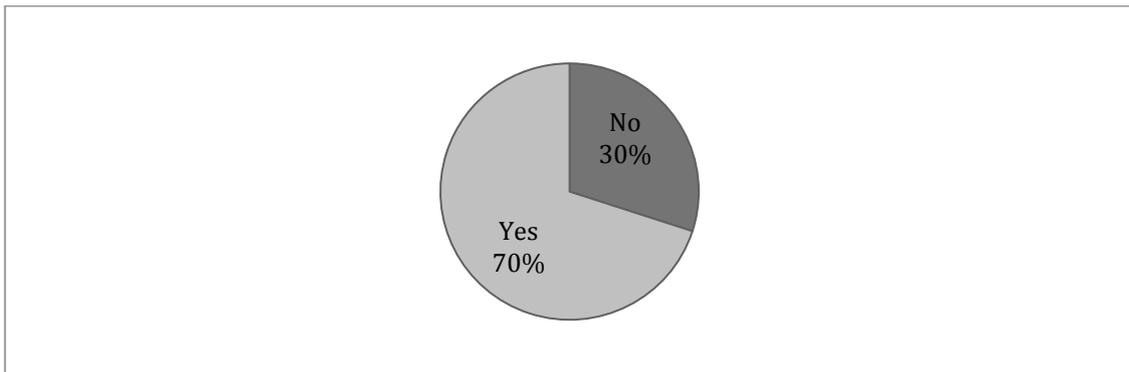
Yes - P1, P3, P6, P7, P10

No – P8, P4

I am not sure - P2, P5, P9



- **Do the images generated by the application inspire your creativity?**



- **(Related to the previous question) why?**

P1 - "I think from the results of the applications the creativity starts"

P3 - "The result was not particularly exciting. on the contrary, the concept is"

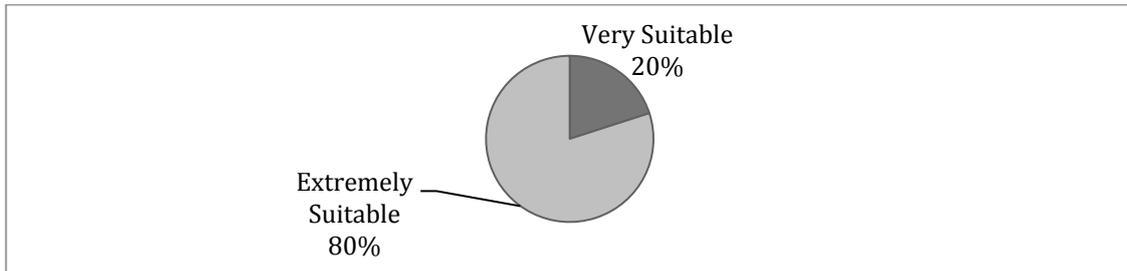
P4 - "Because it gives more possibilities for image editing."

P7 - It's practical and effective; it helps the artist to access the image from other perspective"

P9 – "Because the application highlights certain elements of images"

P10 – 'The app quickly segments images and leave more time to creativity'

- **Do you think the AVC Image Segmentation would be suitable as a plugin for graphics editor software (eg. Adobe Photoshop)?**



Appendix E [Accompanying Folder]

AudioVisual Collage Software

- OSX Application
- Windows Application
- Processing Sketch

AVC Output

- Car.mp4
- Bombings.mp4
- Pop.mp4
- Sea_of_Fog.mp4

Awards

- AES 12Tones Silver Award
- AES Interactive Art Gallery Gold Award
- AES LiveRoom Gold Award

Collaborative Process

- Interactive Art Gallery
- Surround Sounds

Exploratory Studies

- 12 Tones
- LiveRoom