

Dividation – Generative Music Video Editing

Julia Stefan

BSc, MDigDes

Queensland College of Art

Arts, Education and Law

Griffith University

Submitted in fulfilment of the requirements of the degree of
Doctor of Philosophy

May 2015

Abstract

The proliferation of online media has sparked the development of non-linear video forms, which allow for the variation of moving image sequences according to user input or other variables. In order to explore both the technological and artistic possibilities of non-linear video, this thesis introduces the idea of generative editing for music videos, building on previous developments in generative art and metacreation. As a first step, a conceptual framework was defined comprising investigations into artistic movements in non-linear video, media studies on generative and interactive art, and an analysis of music video. These concepts served to support the subsequent development of algorithmic approaches to music video editing and their implementation and testing in *Dividation*, a prototype software system utilised to explore generative music video editing methods using the author's original artwork.

While most non-linear videos tend to emphasise user interactivity, the development of *Dividation* aimed to derive generative methods capable of creating a different sequence every time the video is viewed, without the need for user input. For this purpose, the role of automation and randomness as means to generate variability in a renewed creative process was closely examined. In addition, the possibility of formulating editing dynamics was analysed, and the potential of this creative practice was explored.

Music video analyses conducted as part of this study have shown that generative editing methods need to be part of an open system allowing editors to be involved in the specification of individual progressive requirements for any given music video. While the possibility of videographic notation or any other future systematic description of editing practice has been considered, the results have also shown that editing practice is currently most successful when relying on substantial human input because music videos depend on a large number of parameters describing both musical and visual features and their changing interplay in these sequences. With these ideas in mind, the development of the generative editing methods for *Dividation* was based on a bottom-up approach, with the author's editing practice and experience guiding the process.

In order to generate the music videos, clip progression was defined according to shot structure, resulting in the highlighting of visual characteristics in individual shots and their progressive relationships. The generative editing methods were based on stochastic decision

processes using Markov chains and served to provide simple narrative structures and audio-visual rhythm. These basic methods were supported by additional temporal structural methods which allowed for both an increased control from the editor's point of view and the implementation of greater narrative detail. The development of *Dividation* was further driven by goals such as accuracy towards predefined structures, modularity, the intuitive practice and understanding of processes, as well as technical factors such as streaming quality.

Surveys were conducted among both audiences and editors to reveal the areas in which the development of the editing algorithms succeeded, and to identify other areas still requiring improvement as a basis for subsequent iterations and future research. While substantial efforts have been made regarding the technical development of *Dividation*, this first exploration of the emerging field of generative music video editing, as defined in this thesis, also highlighted the need for further research on the relationship between the creative practitioner and his or her tools with regard to the process as well as the outcomes.

Compared to traditional editing, generative editing is a more rational process, which relies on the editor's ability to translate visual outcomes into values then used by the algorithm to produce these outcomes. Generative music video editing therefore becomes a practice partly removing the author from the process by reducing the importance of manual execution. At the same time, the editor becomes more deeply involved because he or she is required to engage with the video sequences on a different level. The editor, to some extent, becomes a teacher, and the computer, in turn, a powerful tool which may support the artist in the discovery of new possibilities.

To read this thesis online, go to www.juliastefan.com/dividation.

Statement of Originality

This work has not previously been submitted for a degree or diploma in any university. To the best of my knowledge and belief, the thesis contains no material previously published or written by another person except where due reference is made in the thesis itself.

(Signed)_____

Name of Student

Table of Contents

Abstract	1
Statement of Originality	3
Table of Contents	4
List of Figures	6
List of Tables	10
List of Original Video Demos and Artworks	11
1 Introduction	12
1.1 Non-linear Video	14
1.2 Objectives	15
1.3 Research and Thesis Overview	18
2 Media, Technology and Video	21
2.1 Video in Online Media	24
2.2 Non-linear Video and Variability	27
2.3 Automation in Media Technologies	33
2.4 Automation and Creativity	37
3 Creative Algorithms	40
3.1 Generative Artworks	43
3.2 The Role of Randomness	46
3.3 Generative Sequencing of Narratives and Videos	49
4 Music Videos	52
4.1 Music Video on TV	55
4.2 Online Music Videos	58
4.3 The Future of Music Video	60
5 Online Music Video Analysis	64
5.1 Structure and Narrative	67
5.2 Linear Online Music Videos	73
5.2.1 Gangnam Style	80
5.2.2 Call Me Maybe	86
5.2.3 Big Jet Plane	90
5.3 Interactive Music Videos	93
5.3.1 The Wilderness Downtown	98
5.3.2 Like a Rolling Stone	103
5.3.3 Do Not Touch	111
5.3.4 Discussion	116

5.4	Music Video Categorisation	118
5.4.1	Music Video Editing Characteristics	121
5.4.2	Music Genre	125
5.4.3	Types of Video Content	130
5.4.4	Levels of Narrative in Music Video	136
5.4.5	Discussion	145
6	Music Video Editing and Structure	148
6.1	Montage Theory	152
6.2	The Shot as Granule	154
7	Dividation: Generative Music Video Editing Algorithms	157
7.1	Background	157
7.1.1	Requirements	157
7.1.2	Technical Realisation	161
7.2	Generative Editing Development	166
7.2.1	Version One	166
7.2.2	Audience Survey	178
7.2.3	Version Two	187
7.2.4	Version Three	204
7.2.5	Author Survey	214
8	Conclusion	221
8.1	The Evolution of Audio-visual Media	222
8.2	The Development of Dividation	223
8.3	The Definition of a New Creative Practice	224
8.4	Future Research	227
References		231
APPENDIX A		242
APPENDIX B		248
APPENDIX C		262
APPENDIX D		281

List of Figures

- Figure 2.2.1: Interactive music video for Ink by Coldplay (Coldplay. Ink, <http://www.coldplay.com/ink/>). 29
- Figure 2.2.2: Screenshots of the music video for *Like A Rolling Stone* (left)(Bob Dylan. *Like a Rolling Stone*, <http://video.bobdylan.com/desktop.html>) and *Love is the Answer* (right)(Interlude Blog. *The Making of Aloe Blacc's "Love is the Answer"*, <http://blog.interlude.fm/?p=240>). 30
- Figure 2.2.3: Interfaces for interactive videos by Rapt Media for Philips (Rapt Media. *Philips' Goal - Mobile Interactive Video Engagement*, <http://www.raptmedia.com/project/philips-designed-play-drives-mobile-video-engagement>.), Maybelline (Rapt Media. *Maybelline New York's goal - highly engaging interactive tutorial*, <http://www.raptmedia.com/project/maybelline-new-york-engagement-conversions>.) and Learning Care Group (Rapt Media. *Learning Care Group's goal - simplify complex information*, <http://www.raptmedia.com/project/learning-care-group-interactive-healthcare-training-video>). 31
- Figure 2.2.4: Interfaces for *A hunter shoots a bear* (tippexperience. NSFW. *A hunter shoots a bear!*, <http://www.youtube.com/watch?v=4ba1BqJ4S2M>) and *Subservient Chicken* (J1428753. *Subservient chicken*, <https://www.youtube.com/watch?v=K0KjqQZquys>). 32
- Figure 2.3.1: *A Grunt's Life*, by Damon Winter, 2010 (PoYi. Third Place, http://www.po yi.org/68/17/third_01.php.) 35
- Figure 5.2.1: Examples for snowclone captions for *Call Me Maybe* (Know Your Meme. *Call Me Maybe*, <http://knowyourmeme.com/memes/call-me-maybe>). 74
- Figure 5.2.2: Screenshots of the Gangnam Style music video (officialPsy, "PSY - GANGNAM STYLE (강남스타일) M/V, <https://www.youtube.com/watch?v=9bZkp7q19f0>). 81
- Figure 5.2.3: Celebrity cameos in "Gangnam Style" (officialPsy, "PSY - GANGNAM STYLE (□□□□□) M/V, <https://www.youtube.com/watch?v=9bZkp7q19f0>). 83
- Figure 5.2.4: The invisible horse dance (officialPsy, "PSY - GANGNAM STYLE (□□□□□) M/V, <https://www.youtube.com/watch?v=9bZkp7q19f0>). 84
- Figure 5.2.5: Justin Bieber, Twitter post (Justin Bieber, *Twitter post*, <https://twitter.com/justinbieber/status/152843702790914050>). 86
- Figure 5.2.6: Screenshots of the *Call Me Maybe* music video (Carly Rae Jepsen VEVO, *Carly Rae Jepsen - Call Me Maybe*, <https://www.youtube.com/watch?v=fWNaR-rxAic>). 88

Figure 5.2.7: Shot juxtaposition using wind as connecting element (NettwerkMusic, Angus and Julia Stone - Big Jet Plane [Official Music Video], https://www.youtube.com/watch?v=yFTvbcNhEgc).	91
Figure 5.2.8: Screenshots of the <i>Big Jet Plane</i> music video (NettwerkMusic, Angus and Julia Stone - Big Jet Plane [Official Music Video], https://www.youtube.com/watch?v=yFTvbcNhEgc).	92
Figure 5.3.1: Screenshots of the <i>The Wilderness Downtown</i> music video (The Wilderness Downtown, http://www.thewildernessdowntown.com).	98
Figure 5.3.2: Screenshots showing three different channels of the <i>Like a Rolling Stone</i> interactive music video (Bob Dylan, Like a Rolling Stone, http://video.bobdylan.com/).	103
Figure 5.3.3: Screenshots of the <i>Do Not Touch</i> interactive video (Do Not Touch, http://donottouch.org/).	112
Figure 5.4.1: Schematic illustration of the additive segmentation of video content.	123
Figure 5.4.2: Schematic illustration of the substitutive segmentation of video content.	123
Figure 5.4.3: Adherence to the song's form by genre.	127
Figure 5.4.4: Segmentation by genre.	127
Figure 5.4.5: Lip-synching by genre.	127
Figure 5.4.6: Linearity by genre.	128
Figure 5.4.7: Lyric references by genre.	128
Figure 5.4.8: Rhythmic editing by genre.	129
Figure 5.4.9: Shots per minute by genre.	130
Figure 5.4.10: Rhythmic editing by content types.	131
Figure 5.4.11: Lip-synching by content type.	131
Figure 5.4.12: Segmentation by content type.	132
Figure 5.4.13: Shots per minute by content type.	132
Figure 5.4.14: Linearity by content type.	133
Figure 5.4.15: Adherence to the song's form by content type.	134
Figure 5.4.16: Lyric references by content type.	134
Figure 5.4.17: Linearity by narrative level.	138
Figure 5.4.18: Lip-synching by narrative level.	138
Figure 5.4.19: Segmentation by narrative level.	139
Figure 5.4.20: Shots per minute by narrative level.	139
Figure 5.4.21: Adherence to the song's form by narrative level.	143
Figure 5.4.22: Rhythmic editing by narrative level.	144

Figure 5.4.23: Lyric references by narrative level.	144
Figure 7.1.1: Example of visual characteristics in a music video shot.	159
Figure 7.1.2: Illustration of the real-time assembly of music video sequences.	164
Figure 7.1.3: Illustration of the planned assembly of music video sequences.	165
Figure 7.2.1: Examples of field sizes in the Big Jet Plane music video (NettwerkMusic, “Angus and Julia Stone - Big Jet Plane [Official Music Video]”, https://www.youtube.com/watch?v=yFTvbcNhEgc).	168
Figure 7.2.2: Locations shown in the <i>Big Jet Plane</i> music video (NettwerkMusic, “Angus and Julia Stone – Big Jet Plane [Official Music Video]”, https://www.youtube.com/watch?v=yFTvbcNhEgc).	168
Figure 7.2.3: Object shots in the <i>Big Jet Plane</i> video (NettwerkMusic, “Angus and Julia Stone - Big Jet Plane [Official Music Video]”, https://www.youtube.com/watch?v=yFTvbcNhEgc).	169
Figure 7.2.4: Combined parameter model created using the three visual parameters.	170
Figure 7.2.5: Shot decision algorithm describing the shot decision process.	171
Figure 7.2.6: Segmented transition model split into the video's three sequences.	172
Figure 7.2.7: Shot transitions according to the individual parameters with reduced complexity.	174
Figure 7.2.8: Shot decision algorithm for a single shot based on the abstract parameter model.	175
Figure 7.2.9: Algorithm for a single shot decision for <i>Dividation version one</i> .	176
Figure 7.2.10: Algorithm for a single shot decision for <i>Dividation version one</i> .	179
Figure 7.2.11: Percentage of participants by age for <i>Gangnam Style</i> .	179
Figure 7.2.12: Participants' viewing habits for <i>Gangnam Style</i> .	180
Figure 7.2.13: Survey summary for <i>Gangnam Style</i> .	180
Figure 7.2.14: Results for <i>Gangnam Style</i> : Videos "liked" by the participants (in percent).	181
Figure 7.2.15: Results for <i>Gangnam Style</i> : Videos "disliked" by the participants (in percent).	181
Figure 7.2.16: Percentage of participants by gender for <i>Big Jet Plane</i> .	182
Figure 7.2.17: Percentage of participants by age for <i>Big Jet Plane</i> .	182
Figure 7.2.18: Participant's viewing habits for <i>Big Jet Plane</i> .	182
Figure 7.2.19: Survey summary for <i>Big Jet Plane</i> .	183
Figure 7.2.20: Results for <i>Big Jet Plane</i> : Videos "liked" by the participants (in percent).	183
Figure 7.2.21: Results for <i>Big Jet Plane</i> : Videos "disliked" by the participants (in percent).	184
Figure 7.2.22: Production of Majestic.	188

Figure 7.2.23: Examples of different types of shots in the <i>Majestic</i> music video.	189
Figure 7.2.24: Examples of the progressive states in the <i>Look... The Sun is Rising</i> music video.	190
Figure 7.2.25: First experimental interface for <i>Divination version two</i> .	192
Figure 7.2.26: Conceptual notes written at the beginning of <i>Divination version two</i> .	194
Figure 7.2.27: Example of the progression of a parameter's dynamics described using the segmented binary decision model.	197
Figure 7.2.28: Subject parameter relationships in the <i>Majestic</i> music video.	199
Figure 7.2.29: Subject and context parameter relationships in the <i>Majestic</i> music video.	199
Figure 7.2.30: Cutting algorithm for <i>Dividation version two</i> .	201
Figure 7.2.31: Adjustment of the GoPro camera on set.	207
Figure 7.2.32: The five locations shown in the music video for Eloise.	208
Figure 7.2.33: Illustration of the time-code based structure of the first fifteen seconds of the <i>Eloise</i> music video.	209
Figure 7.2.34: Shot decision algorithm using the time-code based narrative structure.	210
Figure 7.2.35: Interactive options for the <i>Eloise</i> music video.	211
Figure 7.2.36: Illustration of the time-code based structure of the first fifteen seconds of the <i>Eloise</i> music video using five-second segments.	211

List of Tables

Table 5.3.1: Like a Rolling Stones Music Video Channels.	111
Table 7.2.1: Possible parameter combinations for a set of three parameters.	198
Table 7.2.2: Participant overview.	216

List of Original Video Demos and Artworks

Example 1.1: Version one, demo 1	170
Example 1.2: Version one, demo 2	173
Example 1.3: Version one, demo 3	174
Example 2.1: Version two, demo 1	192
Example 2.2: Version two, demo 2	194
Example 2.3: Version two, demo 3	204
Example 3.1: Version three, rendered demo	212
Example 3.2: Version three, live demo	212

1 Introduction

“Soon, the way we approach making films and videotapes will drastically change. The notion of a ‘master’ edit and ‘original’ footage will disappear. Editing will become the writing of a software program that will tell the computer how to arrange ... the information on the disc, playing it back in the specified sequence in real time or allowing the viewer to intervene. ... Different priorities rule how and in what order one lays material down on the ‘master’ (disc). New talents and skills are needed in making programs – this is not editing as we know it.”

(Viola [1982] 2003, 467)

As a child, I was fascinated by listening to my parents and grandparents talking about their youth and the technological developments that have happened during their lifetimes. My mother would tell me that they were among the first families in their street to own a television set, and that all the children in their neighbourhood would come over to watch the adventures of *Kasperl*, a popular Austrian puppet show, which until then could usually only be seen live, at unrecorded performances at special events.

How was it possible that my mother grew up in a world where the television set was not a natural element in everyone's household? As a child this was baffling to me.

I did not realise it back then, but eventually, for me, anecdotes like this one led to an essential understanding: Our lives, and therefore our perception of reality are intrinsically shaped by the technologies we grow up with. The technologies surrounding us influence how we communicate, how we work, and how we create.

But what we grow up with are not just existing technologies, but (perhaps even more so) technological revolutions we anticipate for the future.

The introduction of some technologies has been surprising, while others remain in active development. When we look at the Oculus Rift, for instance, we notice that its overall design still largely adheres to early conceptions of virtual reality from the 1980s and 1990s, and the same problems that still require solutions.

Lunenfeld (2004, 282) has called virtual reality "the technological cousin" of interactive cinema, and suggested that both can, unfortunately, only function "in the realm of myth". This notion is hardly surprising. Much like virtual reality, interactive cinema has never found widespread application (Piacenza et al. 2011).

However, starting in the 1990s, digital and online media have rapidly invaded our lives and began to influence media creation and consumption by turning the computer into a mass medium, resulting in what Manovich (2001, 6) referred to as the "computerisation of culture". According to Manovich (2001, 9), this computerisation has lead to "the emergence of new cultural forms" such as computer games, and a redefinition of established media and creative practices, such as cinema and photography.

The use of software and human-computer interfaces have allowed for an unprecedented amount of control over media access, creation and manipulation (Manovich 2001, Manovich 2013). Therefore, the relationships between audience, media and author have begun to change (Rieser and Zapp 2002, Flew 2008). Increasingly, the audience's choice and individuality is foregrounded in media creation, and online media have provided a suitable playground for experiments based on these novel concepts (Manovich 2001, Jenkins 2006).

The new media allow audiences to become media broadcasters and creators themselves on social networks. Media production subsequently has to adjust itself towards these changing dynamics. Therefore, more and more research is dedicated to analysing online media and audience preferences and behaviour (Berger and Milkman 2012, Shifman 2014). In this context, the online video platform YouTube is of particular interest (Burgess and Green 2009, Snickars and Vonderau 2009, Jenkins, Ford and Green 2013, Vernallis 2013). Online video, especially on YouTube, can be said to have transformed the way we think about audio-visual media (Lotz 2007).

This, however, is hardly an isolated phenomenon. Film and video have seen constant transformation since their introduction in the late nineteenth century. In fact, according to Rodowick (2008), technological evolution and therefore reinvention are essential characteristics of all media. They are "continually in a state of self-transformation" (Rodowick 2008, 42).

Online video, while based on existing audio-visual media from cinema to television, has hardly begun to establish itself as a form. With evolving technologies, an increasing number of possibilities for these media are emerging. This thesis aims to discuss one such possibility: generative music video editing.

The concept of generative music video editing offers, on the one hand, the introduction of a new creative practice. On the other hand, it represents a novel audio-visual format harnessing evolving technologies. Generative music videos are non-linear videos which, rather than being edited once and for all, are assembled anew every time a viewer presses play on the interface.

1.1 Non-linear Video

In this context, what exactly does the term 'non-linearity' refer to?

At this point, it is necessary to introduce a distinction between different types of non-linearity applicable to video content. We differentiate from a film theory, and from a technology perspective.

First, from a film theory point of view, Dancyger (2011) distinguished linear from non-linear forms according to causality and sequence progression with respect to narrative logic in a film or video. In this context, non-linear videos predominantly focus on emotion and visual rhythm rather than on the purposeful depiction of a plot. According to this definition, music videos, for instance, are often non-linear because they tend to depict a combination of performance and other visual elements that do not always display an inherent narrative logic.

Second, from a technology perspective, non-linear video refers to sequences that can be altered to some extent, leading to a variable viewing experience comparable to the concepts of interactive cinema. It seems like the music video format, in particular, seems to be a promising candidate for experiments in non-linearity based on interactivity, user input, and alternative methods of assembling video material.

Because they represent a comparatively new element in online media at the time of this writing, the terminology used to describe these non-linear videos tends to vary. The terms "hypervideo", "interactive video" and "multimedia presentations" are used in literature to refer to such videos in addition to "non-linear video" (Meixner et al. 2014).

While the term “interactive music video”, as well as “interactive video” in general, are becoming more common, they do not describe the concept of non-linear video sequences adequately for the purposes of discussing the generative editing methods described in the following chapters. Non-linear video provides for variability both with or without user input. Interactive video provides the viewer with the ability to directly influence a video sequence. While future research projects might consider user interactivity as an option, this thesis focuses on the development of software algorithms as creative tools for providing variable, non-linear video sequences.

The benefits of creating these variable sequences – with or without interactivity – include the exploration of artistic and technological possibilities and the creation of added value for audiences. The methods developed for this thesis will be referred to as generative music video editing, and the following chapters will discuss the use of this term.

First and foremost, generative editing can be differentiated from traditional editing by the introduction of automation for creative purposes.

Naturally, the practice of editing has been subject to technological transformation over the years, for instance, with the introduction of non-linear editing software (NLE). Here, the term non-linear is used to refer to the mode in which footage is accessed and edited. Previously, editing was characterised by the physical process of cutting footage. NLEs introduced a software alternative, which allowed for non-destructive editing and an increased editing efficiency.

The type of automation involved in generative editing is not based on the aim of improving the traditional editing process. Rather it is used to introduce a form which may exist in parallel to linear film and video.

1.2 Objectives

The topic of automation in editing has received increasing attention in the scientific community since the 1980s. Academic interest in the topic continues to grow with the wide availability of video recording devices in conjunction with the importance of audio-visual sequences in mass media. Research on automation in editing varies widely in scope and

purpose, from artistic experimentation (Manovich and Kratky 2005, Chambel et al. 2007) to efficiency-based tools that facilitate speedy film and video production (Girgensohn et al. 2000, Bocconi and Nack 2004, Wang et al. 2005, Kubicek et al. 2008, Kosmopoulos et al. 2009, de Lima et al. 2012, Nithya and Privadhrsini 2013).

The biggest challenge for this type of research and practice is the fact that film and video editing does not adhere to a grammar, notation, or other rules and frameworks, which would dictate how to edit a sequence for a given material. The majority of editors describe editing practice as a complex and highly intuitive process. According to Sheldon Kahn, editor of *Ghostbusters*, *One Flew Over the Cuckoos Nest*, and *Out of Africa*, for instance, knowing how to edit is "strictly something that's inside [the editor] and it's an emotional thing" (Oldham 1992, 21).

However, this lack of a formal grammar and notation is hardly surprising. Film editing was only established in the 1910s, notably with the works of D. W. Griffith and subsequently the soviet filmmakers such as Sergei Eisenstein and Lev Kuleshov. They were among the first who discussed editing in numerous theoretical works (Gunning 1994, Dancyger 2011). Editing is therefore still a very young practice.

The goal behind the design of a generative music video editing system is to create a method that can provide a high level of variability for non-linear video editing without the need for user interaction during playback. A generative music video may therefore be different every time it is viewed. Accordingly, this research revolves around describing the practice of music video editing in order to communicate the dynamics of this practice to a program, which will then proceed to assemble sequences automatically using constrained randomness as a catalyst for variability.

Again, the choice of terminology has to be carefully considered. The word 'automation' itself tends to elicit negative connotations and associations towards efficiency at the cost of artistry. Efficiency, however, is not a primary focus in this research. Instead, the priorities are directed towards creativity and how to maintain it while including automation in this practice. Thus, technology can provide us with tools that support our creativity, stimulating thought and leading to new ideas (Shneiderman et al. 2006). Because of this strong focus on creativity, the terminology used to discuss the project is borrowed from generative art because generative art

provides us with concepts that foreground the use of rule-based procedures, the creative use of randomness as well as elements of variation and unpredictability.

Previous research on automated editing for music video includes, for instance, sequence creation out of home video footage and sports broadcasts (Foote, Cooper and Grgensohn 2002, Wang et al. 2007). In both cases, the two main goals of applying automation in the video editing process were to decrease labour time and to provide editing methods for consumers without professional knowledge. With more and more consumers who have access to video cameras but may not wish or simply do not know how to edit video footage, the latter is becoming increasingly important.

The design of generative editing methods described in this thesis will focus on professional practice rather than support for consumers because the support is already being covered by a growing amount of research and development.

The thesis will introduce the development of *Dividation*, a generative music video editing system. Based on algorithmic methods inspired by the divination practice of the *I Ching*, this project tried to harness the power of computation for the purpose of exploring creativity with the help of automated processes. This way, the research was aimed to define the basic methods of generative music video editing by synthesising current editing trends, if there were any. While this aspect of imitation is dominant in the present study, it is anticipated that future practice based on these methods may establish medium-specific aesthetics, and that the practice will evolve in its own right rather than for the purpose of transforming existing music videos.

Future platforms for generative music videos might offer a unique and novel way of marketing music to viewers by providing a social network around this idea. Viewers will never see the same video twice but will instead encounter a music videoscape which, for instance, will provide different takes on a video's narrative progression or variations on a performance by showing the musician(s) from different angles or in different locations. Users may then be able to set preferences that prioritise certain features of the music video, e.g., its overall aesthetic parameters such as colours, progressive elements determining the prioritisation of narrative or performance elements, and other variables. Generative music videos may even be altered according to the time of day, the weather, or the season, depending on the filmmaker's creative intentions.

What is required, therefore, is a suitable approach and an associated computational tool that provides assistance for the creation of these generative music videos. More importantly, at this stage a basic mechanism is needed that can (1) synthesise editing dynamics to provide adequate values for generative editing, and (2) assemble video sequences in a way that provides variation yet coherence with regard to both the editor's ideas and basic editing conventions.

1.3 Research and Thesis Overview

At the core of this thesis is the development of *Dividation*, a prototypical system for generative music video editing. This development was based on a bottom-up approach, with the system gradually evolving through interactions with my own creative practice as a music video maker and by developing an understanding of generative music video editing processes and requirements.

The primary goal of this thesis was to design both the concept of generative editing itself as well as a prototype system to support the editor, which instantiates an intuitive approach, thus setting the scene for a future implementation as part of a full-fledged software tool. In this respect, the practical research served to rationalise the processes described above, which will have to become automated in order to enable the computer to assemble the video sequences while, at the same time, maintaining the editor's creative input.

This thesis is concerned with automation in so far as it allows us to utilise new technologies to influence art and design. It will be shown that film and video, and therefore the practice of editing, strongly depend on the creative mind. Therefore, a systematic approach can hardly be applied to editing. The development of *Dividation* therefore stresses the importance of the creative partnership between the editor and the tool.

The process of developing *Dividation* was divided into three phases. First experiments served to define basic processes for sequence creation by recreating an existing music video. In the following phase, original music videos were created specifically for the generative editing experiments, and first attempts at defining a workflow from the planning to the final editing were described. For this purpose the previous processes were refined and essential structuring methods were created defining both the aesthetic outcomes, and the creative practice. The

third phase served to further refine these processes and to examine the shifting of creative responsibility between editor and machine. This development process is described in detail in Chapter 6.

Parallel to the development of *Dividation*, theoretical research was conducted to inform these processes.

The development of *Dividation* was first influenced by the growing importance of video content in online media. The current technological evolution and resulting trends in mass media as well as their influence on the music video format are discussed in Chapter 2.

Following this, Chapter 3 examines the role of automated procedures in art and design. Some generative artworks will be discussed and both the potential and the risks involved in the use of these computational methods will be highlighted. In addition to automation, the concept of randomness constitutes an important aspect which must be considered in the development of generative editing methods. Because some autonomy over the creation process is handed over to the machine, the role and extent of artificial intelligence and computational creativity must be taken into account as well. Chapter 3 will therefore discuss varying levels of artificial intelligence in generative artworks in order to determine how much creative responsibility is passed on from the (human) artist to the machine.

At the beginning of Chapter 4, both music video aesthetics and its context in culture and media will be defined. For this purpose, the history of music video will be briefly outlined as well as its evolution through mass media. Here, some aspects of film theory will be introduced to differentiate music video from classical Hollywood cinema, particularly with regard to editing dynamics, patterns, and conventions.

The characterisation of music video and its editing dynamics continues in Chapter 5 with several analyses, first in the form of individual case studies, and second in a comprehensive analysis of a larger body of music videos. The case studies will serve to both highlight some of the features of popular online music videos and to differentiate linear from non-linear music videos while the second analysis aimed to derive a potential form of music video categorisation.

The conclusion reiterates the central claims of the thesis and outlines topics for potential subsequent research projects.

2 Media, Technology and Video

“We should not be surprised that both trajectories - the development of modern media and the development of computers - begin around the same time. Both media machines and computing machines were absolutely necessary for the functioning of modern mass societies.”

(Manovich 2001, 22)

Technology advances are providing the basis for the ever-changing nature of our media landscape and have sparked immense creativity in dynamic, adaptable and user-centred design, leading to the development of what Manovich (2001) has called “new media objects” – digital media elements that are moving away from linear presentation and pre-determined streams of information. These media objects can be characterised first and foremost as being digital, and subsequently by features such as modularity, variability, automation and networking (Manovich 2001, Lister et al. 2009). The advantages of the development of these new media are an increased efficiency of data storage, a high-speed non-linear access, and improved manipulation methods.

It is easy to think of new media as a replacement for “old” media. This may be true for some digital forms of more traditional media, e.g., digital readers, online video on demand, or digital GPS-driven maps on mobile devices. But the predecessors of new media rarely disappear. Rather, they tend to take on a more specialised role (Manovich 2001, Jenkins 2006).

In this context, a term closely associated with new media is “convergence”. It is used to discuss how “old and new media will interact in ever more complex ways” without necessarily displacing each other (Jenkins 2006, 6). The concept of convergence, with regard to media and technology trends, first became popular after Pool (1983) discussed the idea, and was taken up again several years later by Negroponte (1995), who, in the light of the increasing digitisation of mass media, anticipated an equally increasing audience engagement with these media, in addition to the complex transitions and transformations of various media forms through technology.

Jenkins (2006, 15) argued that convergence “alters the logic by which media industries operate and by which media consumers process news and entertainment”, thereby changing both media production and consumption.

There is a risk in trying to adapt and modify traditional media within their established environments in order to fit into the new media landscape. Most manifestations of interactive TV, for instance, have been guided by the notion that traditional media can be more attractive and engaging to audiences by implementing interactivity in order to influence depicted narratives (Vorderer 2000, Kim 2001, Deery 2003, Boddy 2004). So far, implementations of this concept have not been as successful as some research had anticipated. In fact, the concept has been largely abandoned due to high production costs and viewers' seeming lack of interest and engagement (Piacenza et al 2011). Instead, the implementation of interactivity for video seems to be evolving more organically in online media, where they either display new experimental and sometimes artistic methods or concrete functions that audiences are looking for. Examples include clickable hot spots¹ that reveal additional information or options for buying merchandise and objects visible in a particular scene of a video, as well as transcriptions that are directly mapped to the audio track, in order to provide the option to search for content mentioned in the sequence.²

Lister et al. (2009, 21) suggested that interactivity will be a key characteristic of new media, stating that “where ‘old’ media offered passive consumption new media offer interactivity” and defining it as “a more powerful sense of user engagement with media texts, a more independent relation to sources of knowledge, individualised media use, and greater user choice”. Television, however, has been largely referred to as a passive pastime (Vorderer 2000, Fiske 1987) and it is therefore questionable if or to what extent audiences want to interact with this established medium. In contrast, viewers are used to interacting with online content where videos are already embedded in interactive interfaces to begin with.

Along with the media themselves, the terms used to discuss them are subject to transformation as well. While in 2001 Manovich provided a detailed analysis of “new media”, in 2013 he acknowledged that the term is problematic. Manovich (2013, 149) suggested that,

¹ The video “Interactive Shop Tour” uses YouTube annotations in order to provide viewers with clickable hot spots. Once a hot spot is clicked a new video is loaded according to the area that was clicked.

I Like To Make Stuff. “Interactive Shop Tour”, *YouTube* video, 6:07. 27 October 2014.
<https://www.youtube.com/watch?v=iWna7WCP-U>

² This feature is used on the e-learning platform Lynda, where each video provides a transcription which facilitates easier navigation.

Lynda. “Homepage”, <http://www.lynda.com>.

“the terms ‘digital media’ and ‘new media’ do not capture very well the uniqueness of the ‘digital revolution’” because all attributes associated with this new state of media derive from the software that is used to create, manipulate and access media objects. “Media”, therefore, “becomes software” (Manovich 2013, 156).

While many scholars have provided definitions of new media (Bolter and Grusin 1999, Wardrip-Fruin and Montfort 2003, Jenkins 2006, Flew 2008, Lister et al. 2009), it may be useful to find a more precise definition for this thesis. The following discussions are less concerned with the wider spectrum of media that may fall into the category of “new media”, but instead focus on “online media”. Artefacts falling into this category will commonly be browser-based or appear in the form of mobile applications. “Online media” constitutes a useful and concise description of the kind of media objects that are central to this thesis, because the following discussions will largely be centred around their online distribution and communication.

The current state of these online media has been influenced by the concept of Web 2.0, which was introduced in 2004. Web 2.0 is centred around an increasing user participation and networking, a browser-based interaction and cross-platform compatibility, as well as an improved semantic linking (Keen 2007, Jenkins, Ford and Green 2013). The development of Web 2.0 can be regarded as a response to the ever-increasing amount of data generated on the Internet every day. Media audiences, their habits and their modes of consumption are also changing because of the ways this data is navigated. Online media are stored in databases, which in turn are accessible via software interfaces. Algorithms are used to analyse this data and “assist audiences in the process of navigating an increasingly complex and fragmented media environment” (Napoli 2014, 345). One essential aspect in online media is the concept of similar or recommended content featured on most video platforms, news outlets, online magazines and other websites. This way, users are able to create, navigate and curate their own streams of information.

This new audience has been described as “media actives”, a term referring to users born since the 1970s and growing up in a world of increasing control over media as well as a rapidly increasing amount of media focused on popular culture (Jenkins 2006, 255, Danesi 2008). Entertainment in particular is seen as the “killer application” for technology and media, and therefore attracts much attention in research and development with regard to user-centred design and viral marketing (Vorderer 2000, Jenkins, Ford and Green 2013).

2.1 Video in Online Media

Video is becoming increasingly ubiquitous in online media. One reason for this development is the use of media for social engagement and networking. For instance, the sharing of video content on social media has become a common source of entertainment and is supported by most videos' short duration. While social engagement as virtual word-of-mouth communication seems to have evolved into the be-all and end-all of online media, in particular from a commercial perspective, it is important to observe where this communication is taking place. While some argue that YouTube has failed in building communities, other social media platforms are emerging and either introduce their own video formats or actively source from other platforms, thereby functioning as an aggregate where engagement over varying media artefacts occurs (Vernallis 2013).

Online video had only just started to become a more common and accessible media element in 2002, when Flash introduced a new video format, which provided a high quality at comparatively small file sizes. Since 2010, when Apple publicly removed support for Flash on their mobile devices,³ the HTML5 standard has been increasingly used for native video implementation without the need for a third-party plugin, such as Flash. At the time of this writing, only a decade has past since its introduction, and viewers have only started to become sufficiently accustomed to online video. In comparison, television was introduced in the 1920s, and finally found widespread use from the 1950s onwards. Therefore, television and its viewing context have been established for more than sixty years.

Online video technology, in contrast, is only in its early stages and requires further development, especially with respect to (1) stable and uninterrupted high-quality streaming, which involves cutting down on file size by using efficient compression algorithms as well as supporting optimal buffering, and (2) the increasing implementation of video as interactive and non-linear content. Further development may therefore in many cases move the video away from its current prevalent position as a separate standalone element towards a common design element on websites and applications blending in with other content.

³ Steve Jobs, "Thoughts on Flash", *Apple*, last modified April 2010, <http://www.apple.com/hotnews/thoughts-on-flash/>.

In early 2013, Vine, a video sharing platform primarily used on mobile devices, was introduced. Vine allows users to record and share videos with a maximum duration of six seconds. These videos can then be embedded on social networking platforms such as Twitter or Facebook.⁴ After acquiring the platform, Twitter heavily promoted Vine, and its instant success was compared to the previous success of the photo sharing application Instagram.⁵ The videos' short duration and resulting small file size, among other aspects, make this kind of video content an ideal means for online social interaction. In July 2013, a comparable video functionality was also added to Instagram.⁶ A similar concept of establishing a creative community around a short video format was adopted by Smule with the app CineBeat, released in December 2012.⁷ Cinebeat assembles recorded video footage into a 15-second music video, and transforms both sonic and visual qualities of the given video footage. The social media platform Facebook is also continuously working on increasing implementation and ubiquity of video functionality.⁸

The implementation of video content as an essential media element of online services and sites can be risky due to potentially large file sizes and the resulting requirement for fast and stable Internet connections on the client-side. However, the potential of video content for marketing, social networking and information sharing is evident and it can be expected that

⁴ Michael Sippey, "Vine: A new way to share video", *Twitter Blog*, last modified January 24, 2013, <https://blog.twitter.com/2013/vine-new-way-share-video>.

⁵ Finlo Rohrer, "Vine: Six things people have learned about six-second video in a week", *BBC News Magazine*, last modified January 31, 2013. <http://www.bbc.co.uk/news/magazine-21267741>.
Chris Taylor, "What Makes Vine So Hot?", *Mashable*, last modified January 30, 2013, <http://mashable.com/2013/01/29/vine/>.

⁶ Kevin Systrom, "Introducing Video on Instagram", *Instagram blog*, last modified June 20, 2013, <http://blog.instagram.com/post/53448889009/video-on-instagram>.

⁷ Dan Fost, "'CineBeat' app by Smule makes music videos", *San Francisco Chronicle*, last modified May 10, 2013, <http://www.sfgate.com/technology/article/CineBeat-app-by-Smule-makes-music-videos-4503906.php>.

⁸ Seth Fiegerman, "Facebook Buys LiveRail to Make Video Ads Better", *Mashable*, last modified 3 July 2014, <http://mashable.com/2014/07/02/facebook-acquires-liverail/>.

Abbruzzese, Jason. "Google Acquires Directr, a Small Business Video Marketing Startup", *Mashable*, last modified 7 August 2014, <http://mashable.com/2014/08/06/google-acquires-directr-small-business-video-marketing/>.

Josh Constine "Facebook Tests Silent Auto-Play For User Videos In Mobile Feed, Foreshadowing Video Ads", *TechCrunch*, last modified 12 September 2013, <http://techcrunch.com/2013/09/12/facebook-tests-silent-auto-play-for-user-videos-in-mobile-feed-foreshadowing-video-ads/>.

Todd Wasserman, "Facebook: Auto-Play Video Ads Will Remain Small Scale This Year", *Mashable*, last modified 24 April 2014, <http://mashable.com/2014/04/23/facebook-video-ads-small-scale/>.

this area of online media will become a dominant subject of research and development. Cisco⁹ estimated that Internet video (excluding file sharing and gaming) will make up 79% of all consumer Internet traffic by 2018. Compared to other online content and services, such as online music, gaming and social networking, between 2012 and 2013 online video experienced the highest growth in the number of times it was accessed.

One reason for this rapid growth is the simple fact that it is now possible for a large and increasing number of people to produce their own creative works, e.g., in writing, images of any kind, and videos (Keen 2007). Furthermore, it is possible to share these creative works with the world- if the world is willing to pay attention. The success of media artefacts seems to depend, first and foremost, on the aspects of social engagement mentioned above. The more people share a given artefact, the more people may subsequently access it. This concept is at the core of viral marketing, which builds on these social aspects and virtual word-of-mouth mechanisms.

On YouTube, the “most viral videos”¹⁰ seem to revolve around humour as well as music. The most watched and shared video on YouTube in 2013 was the music video for the Norwegian band Ylvis’ song *The Fox (What Does the Fox Say?)*.¹¹ The song revolves around the concept of imagining a variety of absurd sounds a fox might make. The video has collected over 450 million views at the time of this writing. A particularly interesting phenomenon in 2013 was the *Harlem Shake*¹², a 30-second video in which a group of teenagers dance to the song *Harlem Shake*. Within 40 days the video was viewed one billion times and has inspired numerous replications. Another popular video in 2013 was Canadian astronaut Chris Hadfield’s cover of David Bowie’s *Space Oddity*¹³, which became known as the first music video from space.

⁹ Cisco. “The Zettabyte Era – Trends and Analysis”, *Cisco*, last modified 10 June 2014, http://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/VNI_Hyperconnectivity_WP.html.

¹⁰ Neha Prakash, “The 40 Most Viral YouTube Videos of 2013”, *Mashable*, last modified 28 December 2013, <http://mashable.com/2013/12/27/viral-videos-2013/>.

¹¹ TVNorge. “Ylvis – The Fox (What Does The Fox Say?) [Official music video HD]”, *YouTube* video, 3:45. 3 September 2013, https://www.youtube.com/watch?v=jofNR_WkoCE.

¹² DizastaMusic. “THE HARLEM SHAKE (ORIGINAL)”, *YouTube* video, 0:35. 2 February 2013, <http://www.youtube.com/watch?v=8vJiSSAMNWw>.

¹³ David Bowiefans. “Astronaut Chris Hadfield Space Oddity song HD, HQ, Major Tom, David Bowie”, *YouTube* video, 5:30. 13 May 2013, <http://www.youtube.com/watch?v=poZCINzxzrQ>.

Edmond (2012) argued that music videos have ideal preconditions to become viral content because music video was conceived in this way before the age of online media, building on the idea of word-of-mouth promotion and repetitive viewing.

2.2 Non-linear Video and Variability

Advancing technologies have offered a chance to question our satisfaction with current ways of media consumption, by providing opportunities for experimentation with new methods, formats, and techniques. Due to cinema and television, audiences are naturally accustomed to linear moving-image sequences. However, online technologies now allow us to move away from this established mode of viewing and to explore what more or what else can be done with moving images on an experimental basis outside the artistic environments of galleries and museums. This development might result in more accessible experimental media forms and a more casual engagement.

McLuhan ([1989] 2004) argued that Western society has prioritised the visual sense over all others, including the auditory sense. In fact, McLuhan ([1989] 2004, 69) suggested that the eye and its visual mode of perception support the idea of events unfolding in sequence, meaning that every event is “in its proper place, at the proper time, and in linear relationship”. If we accept this premise, then something must have changed these perceptive priorities when we look not just at the nature of online media, but also at their increasing cultural dominance. One of the most recent developments in this area is non-linear video, commonly also referred to as interactive video.

Common features of non-linear video include the deliberate use of non-linear structuring, alternative playback paths or narrative strands, and potential implementation of user interactivity (Meixner et al 2014). In the following chapters, non-linear video will refer to any video sequence that allows for some variability in the order or general appearance of events in the sequence. Any video that can change by itself or allow the viewer to influence what is happening is considered as non-linear video for this purpose. This definition includes but does not equate to interactive video.

At the heart of the idea of non-linear video are concepts of variability, as well as adaptability. This thesis especially considers the potential of variable video sequences to become not just a

form essentially native to online media, but in this way also a parallel to already established linear video forms.

The discussions in this thesis will focus on this idea with regard to music video, and will explore how variability can be achieved. Furthermore, different types and levels of variability will be examined.

Many interactive videos focus on user engagement¹⁴ and on providing game-like¹⁵ or gamified¹⁶ experiences. Other videos with a stronger orientation towards narrativity will apply what we may refer to as tree structures or choose-your-own-adventure narratives¹⁷. This type of structure usually consists of a number of different narrative strands, which are combined and alternated according to either the viewer's choice or chance. Most of these videos will therefore provide visual indicators at fixed points during the video, prompting the viewer to make a decision on what he or she would like to see next (cp. Figure 2.2.1). This method has become popular because it is comparatively simple to implement and provides the video director with a certain level of control and predictability over possible sequence outcomes. This amount of control, however, can also be seen as a disadvantage since it limits the extent of possible outcomes to a finite and relatively small amount.

¹⁴ tippexperience. "NSFW. A hunter shoots a bear!", *YouTube* video, 0:30. 25 August 2010, <http://www.youtube.com/watch?v=4ba1BqJ4S2M>.

¹⁵ ROME. "Homepage", <http://ro.me>.

¹⁶ Do Not Touch. "Homepage", <http://donottouch.org/>.

¹⁷ Coldplay. "Ink", <http://www.coldplay.com/ink/>.

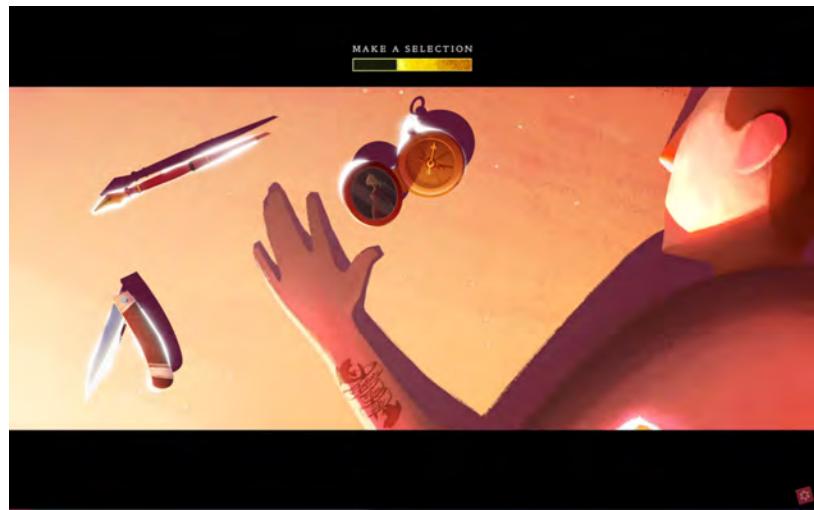


Figure 2.2.1: Interactive music video for Ink by Coldplay ([Coldplay. Ink](http://www.coldplay.com/ink/), <http://www.coldplay.com/ink/>).

In order to provide a truly variable video experience, it is first necessary to find methods that can facilitate the production process for this type of video. Then, the essential method we need to identify must allow to automatically assemble video footage into a sequence.

This is why automation, here, is seen as a necessary step in the creation of these variable videos. Naturally, variability provides several challenges, which may have been the dominant reasons why its development in online video has been comparatively slow. Film and video production is generally a labour-intensive process. Even in their most minimal configurations, production usually requires at least a small team of filmmakers for photography, lighting and direction, as well as actors and actresses or other performers, and eventually editors for the post-production phase. The creation of variable videos would then necessarily result in a multiplication of the required efforts.

Production and post-production times can vary greatly, depending on the budget and overall scope of the film or video produced. Similarly, the additional labour resulting from non-linear video production cannot be quantified – it will depend on the project's concepts and scope.

Aside from the added effort required for the production of footage, the creation or editing of a non-linear or variable sequence from scratch would have to include additional know-how of online and web design technologies and algorithmic composition, as well as time and budget.

Unless sufficient funding is available, this will rarely be a feasible approach. Thus, although interactive music videos have become relatively popular in recent years, they still are not

particularly widespread, demonstrating the need for an appropriate software that can provide some standardised workflows and frameworks. And indeed, we can currently observe the emergence of software companies and media agencies publishing tools to facilitate the creation of interactive videos.

In 2013, the digital media company Interlude¹⁸ was one of the first companies to release an authoring tool for interactive online videos, called *Treehouse*. Interlude was founded by Israeli musician Yoni Bloch who started out by experimenting with interactive music video.

Videos created with *Treehouse* are based on tree-structures. Examples are Bob Dylan's *Like a Rolling Stone*¹⁹ (cp. Chapter 5.3.2) and *Love is the Answer*²⁰ by Aloe Blacc (cp. Figure 2.2.2).



Figure 2.2.2: Screenshots of the music video for *Like A Rolling Stone* (left)(Bob Dylan. *Like a Rolling Stone*, <http://video.bobdylan.com/desktop.html>) and *Love is the Answer* (right)(Interlude Blog. *The Making of Aloe Blacc's "Love is the Answer"*, <http://blog.interlude.fm/?p=240>).

In 2014, Rapt Media²¹ also released an authoring tool for interactive video based on similar structures, but with a stronger focus on gamification and product advertising, rather than music video.

It can be expected that more and more tools for the creation of non-linear videos will become available over the next years.

¹⁸ Interlude. "Homepage", <http://www.interlude.fm>.

¹⁹ Bob Dylan. "Like a Rolling Stone", <http://video.bobdylan.com/desktop.html>.

²⁰ Interlude Blog. "The Making of Aloe Blacc's "Love is the Answer""", *Interlude*, 11 February 2014, <http://blog.interlude.fm/?p=240>.

²¹ Rapt Media. "Homepage", <http://www.raptmedia.com/>.

On the one hand, current trends in non-linear video, particularly in commercials and advertising campaigns, are based on providing entertaining value and addressing a viewer's curiosity with the possibility to explore what these videos have to offer.²² On the other hand, they can provide a customised experience which delivers information²³ relevant to the target audience. Other applications of interactive video include educational purposes and quizzes²⁴ (cp. Figure 2.2.2).



Figure 2.2.3: Interfaces for interactive videos by Rapt Media for Philips (Rapt Media. *Philips' Goal Mobile Interactive Video Engagement*, <http://www.raptmedia.com/project/phillips-designed-play-drives-mobile-video-engagement>), Maybelline (Rapt Media. *Maybelline New York's goal - highly engaging interactive tutorial*, <http://www.raptmedia.com/project/maybelline-new-york-engagement-conversions>) and Learning Care Group (Rapt Media. *Learning Care Group's goal - simplify complex information*, <http://www.raptmedia.com/project/learning-care-group-interactive-healthcare-training-video>).

Over the coming years, when this type of media format becomes more commonplace, methods of using interactive video for advertising will require increasing development and analysis. While some research has addressed viral campaigns such as *Subservient Chicken*²⁵, or Tipp-Ex's *NSFW. A hunter shoots a bear!*²⁶ (Jenkins, Ford and Green 2013, Vernallis 2013, Shifman 2014, Figure 2.2.3), it is hard to predict the success of similar campaigns in the future once users get accustomed to this kind of presentation and advertising methods. An important consideration in the development of non-linear methods for online video is the value of novelty, which in these cases relies on a viewer's curiosity and eagerness to explore the video's capabilities.

²² Rapt Media. "Philips' Goal - Mobile Interactive Video Engagement", <http://www.raptmedia.com/project/phillips-designed-play-drives-mobile-video-engagement>.

²³ Rapt Media. "Maybelline New York's goal - highly engaging interactive tutorial", <http://www.raptmedia.com/project/maybelline-new-york-engagement-conversions>.

²⁴ Rapt Media. "Learning Care Group's goal - simplify complex information", <http://www.raptmedia.com/project/learning-care-group-interactive-healthcare-training-video>.

²⁵ Subservient Chicken. "Chicken Redemption", <http://www.subservientchicken.com/>.

²⁶ tippexperience. "NSFW. A hunter shoots a bear!", YouTube video, 0:30. 12 August 2010, <http://www.youtube.com/watch?v=4ba1BqJ4S2M>.



Figure 2.2.4: Interfaces for *A hunter shoots a bear* (*tippexperience. NSFW. A hunter shoots a bear!*, <http://www.youtube.com/watch?v=4ba1BqJ4S2M>) and *Subservient Chicken* (J1428753. *Subservient chicken*, <https://www.youtube.com/watch?v=K0KjQZquys>.)

An essential question for extending, transforming and developing these methods is: what do users really want? At the time of this writing, TV and online commercials seem to benefit from offering interactive media instead of or in addition to more traditional forms of advertising. However, once the novelty of these formats wears off, viewers may stop engaging with these ads. In many cases, when it comes to advertising, they may not even want to see them in the first place.

Although the development of generative editing methods in this thesis focuses on music video, it is possible that the explored editing methods can provide solutions to these types of problems as well by providing variation without turning the viewer into an active participant.

Without requiring user input, any type of video that is generated with these editing methods can provide variation that, aside from creative considerations, may be more engaging for viewers because the created videos won't repeat themselves, but rather provide a kind of videoscape in which the unfolding events can be viewed from different perceptual angles.

So far, particular areas of interest in the research of non-linear and interactive video tend to revolve, for instance, around automated video composition and editing for the purpose of interactive storytelling (de Lima et al. 2012), creating video art and exploring evolutionary algorithms (Chambel et al. 2007), personalisation of video artefacts (Kosmopoulos et al. 2009) and hypervideo systems (Leggett and Bilda 2008). An increasing number of researchers are also looking into the automation of manual video editing processes in order to create products more appealing to consumers and to increase the efficiency of various production procedures (Foote, Cooper and Girgensohn 2002, Hua, Lie and Zhang 2004, Kubicek et al. 2008, Chen, Delannay and de Vleeschouwer 2011).

This thesis and the practical experiments discussed in Chapter 6 will focus on the development of new techniques in online video production in order to (1) enhance a viewer's experience by providing videos that alter every time they are accessed, and (2) explore new creative methods and potential tools for video creation.

2.3 Automation in Media Technologies

Advancements in technology originate in the desire to improve a certain functionality. But such improvements cannot happen entirely without disruptions. Not surprisingly, in this context, automation tends to be a term causing apprehension, primarily because of robots and other technologies, which increasingly perform tasks that once were covered entirely by manual labour.

While advancing technologies may enhance the potential of creative people, Shneiderman (2000, 119) argued, methods that automate creative processes will promote widespread creativity and therefore raise "expectations that may change employment patterns, educational systems, and community norms". Furthermore, he suggested that "introducing computer supports for creativity may produce greater social inequality as it raises the costs for those who wish to participate". Noble (1998, 368) studied the role of digital technology in education and argued that so-called "digital diploma mills" are threatening the quality of higher education. Hu, Clark and Ma (2003) examined the acceptance of technology in teaching and highlighted some of the issues for teachers who have to adopt new tools into their routines, such as perceived relevance and usefulness as well as computer literacy. Frey and Osborne (2013) discussed the issue of employment with regard to automation and computerisation which has affected social equality and occupational structures. Automation, in this context, led to the disappearance of many traditional occupations, but also introduced new roles for people.

We usually speak of automation when we devise strict rule-systems and procedures that define the course of certain actions and processes. Eventually, technology will often and increasingly be employed to execute these automations, on the one hand, indeed, for the sake of efficiency, but on the other hand also to provide room for a person to dedicate their minds to potentially more useful and creative activities that cannot be fulfilled by a tool (Nye 2006, Frey and Osborne 2013).

Rodowick (2007, 99) defined “the digital arts, if there are any” and new media as characterised by automated procedures and therefore by self-actualising actions that respond to specified commands. Manovich (2001) further asserted that the use of automation in new media facilitates the organisation of digital data and can further be used to remove human intentionality to a certain extent.

In this context, photography, “the most important event in the history of the plastic arts” (Bazin and Gray 1960, 9), will serve as a useful example of recent developments in art and technology.

With the *camera obscura*, the idea of photography first manifested itself as a technique to research and reproduce perspective. According to Daval (1982, 12), it was “a direct consequence of the evolution of drawing”, which had until then been “the only way of recording observations and visual knowledge”. When chemical photography was finally invented in the nineteenth century, it found application in various domains, from portraiture to scientific inquiry into motion, early press photography, and as a method of replication.

Some painters, such as Baudelaire, saw photography as a branch of industry and therefore refused to accept the technology as an art form. According to Daval (1982, 83), painters considered photography as “a threat to their livelihood” since they feared that it could “provide those delineations of reality which they alone had provided”. Those painters weren't necessarily wrong in their fears. However, whereas Paul Delaroche allegedly announced that painting had died in the wake of photography (Crimp 1981), Bazin declared that photography as “a liberation and an accomplishment … has freed Western painting, once and for all, from its obsession with realism and allowed it to recover its aesthetic autonomy” (Bazin and Gray 1960, 9).

In what can be considered an ironic turn of events, photographers would hold a similar position as these early painters when, in 1888, George Eastman invented the Kodak camera, which allowed a broad audience to record photographs with a device that was easy to use, and a service that would develop the recorded photographs for them (Daval 1982). Until then, photography had only been practised by a comparatively few number of researchers and professionals, since it was a laborious process that required technological expertise.

It should not be surprising that the death of several artistic practices due to technological developments has been predicted every so often. At the end of the 1980s, the emergence of Photoshop was seen as a detriment to the art of photography, in particular to photojournalism (Keith 2014). Now, photography is once again perceived to be in mortal danger, as mobile devices such as smartphones have provided virtually everyone in the industrialised countries with a camera that accompanies their every step. “Iphoneography”, according to Halpern and Humphreys (2014, 17), “represents an example of an emerging art world currently in the process of legitimation by distinguishing the process, artefacts, and actors from mass consumers of iPhones and photo apps”. An example of photojournalism created with a mobile phone that has been discussed in this context is *A Grunt's Life* by Damon Winter (cp. Figure 2.3.1), a photo series which has won third place at the Picture Of the Year international (POYi) contest in 2010 (Alper 2013, Halpern and Humphreys 2014, Keith 2014). The controversial discussions centred around the use of photo apps such as Instagram revolve around the application of filters and increasing anaesthetisation because of questions regarding authenticity and appropriateness, especially in photojournalism. Additionally, according to Keith (2014, 1240), “anxiety derives from the fear that not only does Instagram make photos look better, but it also makes anyone's photos look better, even potentially better than the photojournalists who have resisted the transition from traditional film stock photography to digital photography altogether”.



Figure 2.3.1: A Grunt's Life, by Damon Winter, 2010 (PoYi. Third Place, http://www.pozi.org/68/17/third_01.php.)

Described by Rodowick (2008, 47) as “the conversion of light into matter as the result of several automatic processes”, photography appears to have been constantly evolving since its invention in the early nineteenth century. In photography, automation has been applied to allow a wider range of consumers to use the technology.

With the use of non-linear editing systems (NLEs), the practice of film and video editing, in comparison, has also seen a significant change in order to allow for a more efficient editing process. These new editing software programs have allowed for the editing process to become more precise, less prone to error, and facilitated an easier management and organisation of footage. Few experts would deny these advantages, but many professionals who started with manual editing have, at least in the early stages, found this renewed approach less enjoyable and perhaps detrimental to their creativity (Oldham 1992).

According to Dancyger (2011, 392) “there is no question that nonlinear editing and digital technology will have positive effects on the editing process and on the outcome of that process, the screen story”. Dancyger further elaborated that “the capacity to work in a more complex way with sound and picture can only help the postproduction process and budget”. NLEs will therefore “enable creative decisions to be arrived at more quickly than earlier editing technology, but will not make the creative decision”. The physical interaction with the medium of film was removed from the editing process by replacing the literal cutting with the figurative cutting of footage. Expertise required to operate the Moviola, for instance, was replaced with know-how of its software equivalent.

Following Cavell (1979), Rodowick (2007, 43) defined media as being dependent on the automatisms of an art form, stating that “no act of creation occurs in perfect freedom”. Cavell (1979, 103) previously elaborated that photography and film have a “natural relation to [their] traditions of automatism”, largely due to their physical mechanisms and technological context. According to Cavell, creating new media for their art is the task of modern artists, which he further defined as “the task of establishing a new automatism” (Cavell 1979, 104). Cavell's (1979, 105) definition of automatisms refers to “discoveries of form and genre and type and technique”. Cavell (1979, 107) explained that his “impulse to speak of an ‘automatism’ is ... due first to the sense that when such a medium is discovered, it generates new instances: not merely makes them possible, but calls for them, as if to attest that what has been discovered is indeed something more than a single work could convey”.

Another motivation for his reference to automatism was to first draw awareness to genre features which may be classified in this way in order to provide freedom to the artist and autonomy for the artwork (Cavell 1979, 107). Rodowick (2007, 44) also elaborated that “in encountering automatisms as limits, artists invent new creative strategies as ways of overcoming or transforming them”.

Art and design can therefore be said to be dependent on their technological context and, rather than defining an art by the physical materials or traditional structural methods used, they should be defined by “the forms of expressiveness” that they enable (Rodowick 2007, 45). A medium, therefore, “is nothing more nor less than a set of potentialities from which creative acts may unfold” (Rodowick 2007, 85).

The idea of automatisms, then, implies that in some ways we already depend on methods that do some of the work for us, if only conceptually. While computational automation is several steps away from Cavell’s automatisms, its notion may seem less foreign or unnatural to the domains of art and design from this perspective. In fact, it appears natural that we would devise methods that prevent us from having to reinvent the wheel, over and over again, when we instead could focus on areas which have yet to be explored.

2.4 Automation and Creativity

Nake (2012, 81) described “the art of writing computer programs” as a “skill of utmost creativity, intuition, constructive precision, and secrets of the trade”. He further emphasised that, while the impact of using algorithmic methods for creative purposes is vastly different compared to other tools, e.g., the painter’s brush, creativity still resides within the person applying these algorithms, not in the machine. The use of algorithms for creative outcomes leads to a more complex relationship between artist and tools and therefore requires a different way of thinking, which, according to Nake (2012, 93), is characterised by “explicitness, computability, distance, decontextualising and semioticity”. Candy and Edmonds (2002, 29) further stated that these cognitive transformations in creative practice caused by more complex interactions with computational systems as creative tools lead to “new questions and new ideas”.

By automating certain tasks, creativity is shifted from their manual execution to more complex activities and functions of control over these tasks. But where does automation begin?

Galanter (2008) highlighted the idea of highly ordered systems in creative practice and suggested that symmetry may be one of the oldest and most used autonomous systems applied in artworks. According to Galanter, a prominent example is the art of tiling with girih

patterns in Islamic art and architecture, which are based on strict mathematical systems (Lu and Steinhardt 2007). Girih patterns have continued to inspire artists from other domains, e.g., M. C. Escher, and Galanter (2008, 318) referred to Escher's methods as "algorithms in the service of art".

However, as Nake (2012) implied, the complete removal of manual execution of certain tasks necessarily introduces a new situation for creative practice. Naturally, it will make us question the essence of art and human creativity. Not unlike the painters who feared their professional purpose may be taken away by photography, as well as the photographers who feared the same once photography became widely available thanks to the Kodak camera, many artists and creative people will remain apprehensive of automated techniques that might perform certain tasks better and more efficiently than they were able to, and thereby potentially threaten their creativity. In his introduction to *The Graphic Work of M. C. Escher* Terpstra wrote:

"Anyone who from early youth has applied himself to the practice of graphic techniques may arrive at the stage where he begins to regard complete mastery of his craft as the highest ideal. The finesse of craftsmanship takes up all his time and demands his attention so completely that often he will make even the choice of his subject subordinate to and dependent on his desire to explore a certain facet of the technique. There is indeed deep satisfaction in acquiring artistic skill, in getting to know intimately the properties of the available materials and in learning the effective and controlled use of our instruments – and in the first place of our hands!"

(Terpstra 1961)

Obviously, a deep traditional view exists linking artistic skill to manual execution. The exploration of algorithmic or generative artistic methods, however, does provide increasing potential for the acquisition of "artistic skill", for "getting to know intimately the properties of the available materials" and for "learning the effective and controlled use of our instruments". The mastery of a form of art then increasingly shifts from the hand to the mind. But we also need to remember that this shift does not necessarily result in replacement.

Some generative computational artistic practices already exist side by side with their more traditional forms, e.g., in the case of certain aspects of generative music, such as live coding, an improvised programming practice often used in conjunction with computer music (Sorensen and Brown 2009).

In this context, Brown (2012) ascribed two essential functions to the use of generative methods, both of which lie in the extension of a person's abilities, either (1) for the skilled practitioner who seeks to engage more deeply with their (art)work through the exploration of different tools, or (2) for the inexperienced artist who may require technological assistance to overcome gaps in know-how and ability.

For generative video editing in the context of this thesis, we are not concerned with speeding up the process of editing linear video sequences, but rather with extending the experience of both author and audience. As a creative tool or partner, the generative computational system then serves to provide new creative possibilities for the author and a mode of viewing that exploits current technologies in order to provide variation. In this sense, whereas video has commonly been a static artefact (outside the field of video installation art and live visuals or VJing, for instance), it may increasingly adopt elements of performance and improvisation.

Depending on the director's creative intentions, a video may contain a large number of different narrative directions and perspectives as well as varied emphases on particular visual characteristics. The video editing process could furthermore include natural variables such as the time of day or the weather, in addition to or as a substitution of computational random values.

3 Creative Algorithms

"Today some media are already totally automated while in other cases this automation hardly exists - even though technologically it can be easily implemented. Let us take as the example contemporary Hollywood film production. Logically we could have expected something like the following scenario. An individual viewer receives a customized version of the film that takes into account her/his previous viewing preferences, current preferences, and marketing profile. The film is completely assembled on the fly by AI software using pre-defined script schemas. The software also generates, again on the fly, characters, dialog, and sets ... that are taken from a massive 'assets' database."

(Manovich 2003, 18)

Harnessing the power of new technologies allows for the automation of tasks that can be easily modelled by algorithms in order to provide room for a more complex expression of the creative mind by shifting creative efforts further towards conceptual tasks and thereby away from manual execution.

Several terms are currently emerging from the discussion, allowing to structure and facilitate the discourse on these novel artistic forms and methods of media creation. Depending on the degree of autonomy exercised by a computational system and its agency, we may speak of metacreation, generative creation, or procedural creation. Frequently, the choice of terminology also depends on the domain in which the process and outcome is located. The term metacreation usually refers to software that "exhibits behaviors that would be considered creative if performed by humans" and is therefore closely correlated to AI (Whitelaw 2004, Eigenfeldt et al. 2014, 321). Procedural methods will often refer to processes designed for performing repetitive tasks not feasible for manual execution, particularly in games and special effects creation, whereas generative methods often describe systems designed for the creation of art.²⁷ All of these terms are used to inform our understanding of creative processes and emergent structures, as well as to extend and transform our creative ability. In the fields of art and design, this is commonly achieved by utilising automation and computational systems.

²⁷ The term "generative" is also frequently used in linguistics, where generative grammar theories are used to explore structure and formal descriptions of language (Chomsky 1975).

This concept of enhancing people's skills and abilities with the help of technology received particular interest in the 1950s, when fields such as cybernetics and AI achieved widespread awareness. In 1960, Licklider (2003) described what he called the “man-computer symbiosis”, and, in 1962, Engelbart (2003) spoke of “augmenting [the] human intellect”, while Nelson promoted his vision of “Thinkertoys”, i.e., a system designed “to help people think”, or more specifically “a computer display system that helps you envision complex alternatives” (Nelson 1987, 50). In the 1970s, Helvey (1971, 4) mentioned the “machine partner” and the necessity of a “cybernetic pedagogy” that would allow human beings to adequately live according to their times, arguing that the study of cybernetics is essential for providing the necessary means of communication between man and machine. Similarly, Ascott ([1964] 2003, 129) suggested that the artist should turn to science, and in particular to cybernetics, in order to “orientate himself in the modern world”.

This development consequently affected engineering, media and art, resulting in the emergence of several art movements, e.g., conceptual art, art-and-technology, computer art, algorithmic art and generative art. All of these new artistic movements shared a common approach focusing on the idea and concept. In conceptual art, this resulted in a lesser focus on the outcome itself, whereas other movements emphasised technological exploration (LeWitt 1967, Shanken 2002, Nake 2010).

Derived from the Greek words “techne” and “logos”, the term “technology” first referred to the systematic study of the arts and associated technical skills. Today, it is widely used to refer to complex systems, machines and computers (Nye 2006).

The increasing use of technology and the ongoing development associated with its introduction are observed with both hope and apprehension. Weizenbaum (1976) argued that the influence of computer technology on society should be subject to certain limits. In particular, he urged that human functions, such as “respect, understanding, and love”, should not be substituted by technology, and also demanded that the development of technology should be carefully considered wherever it may have “irreversible and not entirely foreseeable side effects” (Weizenbaum 1976, 270). Other critics are pointing to a general decline in artistic skill and technique over the reverence of idea and concept (O’Hear 1995). For instance, O’Hear argues that technology is removing constraints on imagination and possibility, leading to the degeneration of craft. Dorin (2013, 3) mentioned “missed opportunities” addressing the controlled environment of digital systems and their resulting lack of natural chance outcomes.

Referring to “systemic films”, Le Grice (1977, 129) argued that some critics consider these systems to eliminate choice. However, according to Le Grice, choice is simply shifted to another area. Eno (2004) also supported this idea, emphasising that, in music composition, technology has not removed constraints or improved the music itself, but has helped to create different kinds of work by shifting constraints.

Stating that it “is easy to imagine human beings as pre-literate, but it is difficult to imagine them as pre-technological”, Nye (2006, 5) described technology as inherently human by drawing on the use of tools to shape and adapt the world. Whereas Weizenbaum (1976, 259) seemed to fear “technological inevitability”, Nye argued against a predictable and deterministic view of technology, instead claiming that culture and social developments are shaping the technologies in use, not the other way around. He suggested that technology may lead to greater choice, possibility, and diversity in society. Societies’ increasing dependency on technology must therefore not be considered a matter of determinism, but a “technological momentum” (210). This momentum increasingly revolves around the computer, which, according to McCormack and d’Inverno (2012, VIII), has transcended the role of a tool to become a “dynamic cultural and creative partner”. Naturally, some artists and scholars support the notion of computers as partners in the creative process, whereas others reject this attribution by referring to the computer as “an automation”, rather than a partner (Brown 2012, Nake 2012, 73). Again, this discussion largely refers to the level of agency of a computational system, as well as the type of interaction required to produce an outcome.

Manovich (2003) has argued that, in our contemporary society, the computer is generally seen in the role of a “technician” (18). In contrast to this still largely prevalent image of the computer, Brown (2012) suggests that we “should be concerned with partnerships with generative processes where the ‘system’ includes a human participant, as opposed to a system as a tool or an autonomous machine”. Further explorations into these creative partnerships should be made to promote the use of emerging technologies and generative procedures in order to advance our understanding of creative practice. Some successful examples of computational creativity can be referred to as “human-system collaborations” (Bown 2012, 374). Popular examples are *AARON* by Harold Cohen and *Voyager* by George Lewis.

Regardless of terminological considerations and different implementations of intelligence and autonomy in a system, algorithms and automation are increasingly used to decrease manual execution while, at the same time, emphasising conception. Following Heidegger ([1949]

1977), McLuhan ([1964] 2001, 7) referred to technology as extensions of man, describing this process as creating “roles for people, which is to say depth of involvement in their work and human association that our preceding mechanical technology had destroyed”. Clark also elaborated on this subject, claiming that the unusual plasticity of the human brain lead to an extended use of tools and technology and a preference for “meta-knowledge over baseline knowledge” (Clark 2003, 67). Although there is a common fear of decline in human technical and artistic skills with increasing use of technology, there is also the potential for creative people to benefit from evolving technologies and to explore new possibilities and domains (Shneiderman 2000).

3.1 Generative Artworks

Although the terminology and its respective domains of application overlap, the following section focuses on the use of the term “generative,” since it provides a description for a wide range of different works in art and design, and may be used to refer to varying levels of autonomy shared between the human author and a respective computational system. Furthermore, it has become an established denomination in the arts, commonly used since 1965, when Georg Nees first presented *Generative Computergrafik* (Boden and Edmonds 2009).

Technically, the history of generative procedures precedes the age of computers. Such procedures were particularly applied in the field of musical composition, with its notation system providing for an extensive set of rules and conventions facilitating the utilisation of algorithmic methods of music creation (Nierhaus 2009, Eigenfeldt et al. 2014). According to Nierhaus, algorithmic musical composition can be dated back to Guido of Arezzo, who not only contributed to the development of musical notation, but was also one of the first music theorists proposing a system for the automatic generation of melodies in AD 1000. Another well-known, albeit much later example of algorithmic procedures in musical composition is known as *Musikalisches Würfelspiel*, a game popular throughout the eighteenth century in Europe, in which the throw of dice determines a musical sequence assembled from pre-composed bits and pieces. The earliest known example was created by Johann Kirnberger in 1757 (Nierhaus 2009).

In the 1960s and 1970s, after both the rising interest in systems theory, AI, chaos theory, etc., and the increasing availability of computer technology, the field of generative art emerged

rapidly (Welsby 2006, Dorin 2013, McCormack et al. 2014). Early artists who established the field include Roy Ascott, Harold Cohen and, of course, Georg Nees, whose PhD thesis on computer art greatly influenced a growing community of computer artists (Boden and Edmonds 2009).

Boden and Edmonds (2009, 29) defined generative art as artworks that are “generated, at least in part, by some process that is not under the artist’s direct control”. According to Galanter (2003, 4), “[g]enerative art refers to any art practice where the artist uses a system, such as a set of natural language rules, a computer program, a machine, or other procedural invention, which is set into motion with some degree of autonomy contributing to or resulting in a completed work of art”. Extending on broad definitions of generative art, McCormack et al. (2014) elaborated that the extent of autonomy practised by the computational system may vary greatly between projects where the artist’s influence might be reduced to an absolute minimum, and projects where the artist’s influence and responsibility remains a primary driver and the computer plays a more passive role in the creative process.

The use of generative procedures in an artistic context seems to naturally result in the discussion of role, meaning and purpose of the computer in these processes. McCormack et al. (2014) endeavour an understanding of generative art by proposing and answering several essential questions, ranging from a philosophical discourse on the nature of art and aesthetics to the role of technical parameters, and also include the overall characteristics of generative art.

First and foremost the discussion typically revolves around the potential source of creativity and originality of a computer-generated artwork. Although some consider it proven that computers can exhibit creative behaviour, pointing to generative artistic systems such as AARON, the necessity of (re)defining art and the boundaries in which generative computational art can operate remains a problematic aspect (McCorduck 1991, Eigenfeldt et al. 2014). Contemplations on aesthetics and beauty strongly depend on artistic judgement and cultural influence, and it therefore seems necessary to introduce formal definitions of what we think art should look, sound or feel like. The role of the computer and its artistic potential in these processes may once again require a redefinition, or at least a reconsideration of the (human) artist, a concern that appears to re-emerge whenever a new technology is introduced that transforms manual practices, which was, for instance, also the case in the fields of photography and cinema (cp. Chapter 2.3).

McCormack et al. (2014) proposed that a judgement on the subjective quality of generative artworks must consider the work's engagement in process over outcome, which is where the artistic intent usually resides. A lot of generative practice is done within the fields of design and technology, "outside the precious bubble of the high art world", where the discourse on process and practice may adequately be of comparatively equal or greater interest than the actual outcome (McCormack et al. 2014, 139).

Generative art projects therefore also cover a wide range of media and are currently largely focused on music, visual art, software, architecture, and literature. Some examples revolving around film and video manipulation include *Soft Cinema* by Lev Manovich and Andreas Kratky (2005), *Re:Cycle* by Bizzocchi (2011), and *whiteonwhite:algorithmicnoir*²⁸ by Eve Sussman. All of these projects are based on the use of databases containing audio-visual footage and other assets, which are then assembled according to pre-defined rules and concepts, as well as the utilisation of computational randomness. By using generative composition methods, the respective software systems are able to create a different sequence each time the video is played.

The application of generative methods in the video domain can serve different purposes, as mentioned earlier, by extending the abilities, on one hand, of skilled practitioners seeking to engage more deeply with their artwork, and on the other hand for inexperienced artists who require some technological assistance to overcome gaps in know-how and ability. Concerning the creative output, we can differentiate between the exploration of a narrative and its flexibility, and research focusing on technological possibilities for creating visual time-based media. Thus, generative video projects allow us both to gain a better understanding of how we perceive a narrative logic and different types of structure, and to experiment with these structures. On the other hand, generative editing methods can simply serve the practical purpose of creating a greater variation, e.g., when different videos are repeatedly shown to an audience but at least some novel element has to be included.

²⁸

"whiteonwhite:algorithmicnoir," <http://www.rufuscorporation.com/wowpr.htm>.

3.2 The Role of Randomness

Generative artistic procedures will often rely on a certain degree of randomness or pseudorandomness to not only provide variation but also a sense of imperfection, an element that can “humanise” the creative work (McCormack et al. 2014, 138). In this context, randomness is used to simulate the complexity of nature by filling gaps in our knowledge and understanding of creative processes.

The application of chance in art and design generally prompts the question of source and originality. Some authors may not accept works where chance is a dominant parameter in the creation of an outcome as art at all. In contrast, others view chance as an essential artistic element in general. Dorin (2013) suggested that the influence of chance on art can be traced back to the emergence of the prevalence of symbolic thinking in hominids. Some early iconic artworks, for instance, the *Venus of Tan-Tan*, were most likely salvaged artefacts that resembled a particular figure and were then modified to increase this resemblance (Bednarik 2003). Chance became a more prominent element in Western art in the early twentieth century with movements like Dada and Surrealism (Follett 2011).

Today, to most people, using chance over a rational process of weighing up facts in order to make important decisions would seem counter-intuitive. However, when logic and reason cannot be applied, we seem to be more willing to make use of randomisers, by tossing a coin, rolling a die, etc.

Using chance or randomness in order to make decisions without resorting to human logic and to complement deterministic processes in daily life can be regarded as an age-old ritual. According to Bennett (1998, 9) there are three basic reasons for using randomisers: (1) to provide an unbiased choice or fairness, thereby (2) preventing conflicts, as well as (3) asking for “divine direction”. In fact, the chance outcome was often seen as a manifestation of divine will and therefore as superior to human choice. Additionally, chance outcomes have also been used to introduce irregularity into activities such as hunting, in order to avoid predictable behaviour (Moore 1957).

The notion of divine intervention used to explain the perception of significance in random events has not completely lost its validity even in secular culture, where a pattern or even a destiny is still frequently read into chance outcomes. Mlodinow (2008, 182) argued that this

behaviour is part of our human nature, since our minds evolved to “assimilate data, fill in gaps, and look for patterns”.

Coincidental events may also be regarded as an expression of nature’s creativity itself (Combs and Holland 1990). We may perceive something as beautiful, unable to immediately grasp its origin or underlying structure, and therefore experience a greater appreciation of this event or outcome.

Some authors suggested that randomness is primarily an expression applicable to structures beyond our comprehension, for example in extreme cases, because they can only be referred to within the field of quantum mechanics (Combs and Holland 1990). Paul Kammerer, for instance, a biologist and Lamarckian evolutionist, postulated his theory of seriality in 1919. Kammerer described the propagation of events as cyclical and suggested that similar events attract each other to form a series. Jung drew on Kammerer’s theory when he proposed his concept of synchronicity, which describes acausal connections between events by referring to archetypal processes (Jung 1967). Jung therefore developed a fascination for methods of divination, specifically the *I Ching*.

The *I Ching*, literally the Book of Changes, most popularly provides a means for divination by casting coins or yarrow sticks. Each cast determines the answer to a yes-or-no question represented in writing as a full or broken line and is repeated six times in order to produce two connected trigrams. The principle of duality in this process also lead to the (re-)discovery of the binary numeral system by Leibniz ([1679] 1994) in the seventeenth century. This is one of many reasons why so much importance is ascribed to the *I Ching*.

With regard to stochastic processes in art, the *I Ching* was most famously employed by John Cage to compose his *Music of Changes*. Cage, the American composer who, in the 1950s, made extensive use of chance processes, described the *I Ching* as a means to remove himself from the activity of music- and meaning-production, similar to various other experimental musicians of the same era who used, for instance, tables of random numbers to achieve a similar effect (Griffiths 1981, Follett 2011). Cage intended to draw attention to sound itself, within its own inherent nature, and allow “the world to speak in its own, raucous, dissonant voice” (Follett 2011, 104). “Aesthetic chance”, Follett (2011, 105) argued, renders the elements in a work interchangeable and replaceable, and thereby opens up the potential for “infinity” by making an object “signify much more than itself” (Follett 2011, 106).

From a rather more technical point of view, Goldreich (2008) summarised definitions of randomness according to three different lines of thought. First, randomness can be considered “a lack of information” (Goldreich 2008, 285). Second, following mathematicians and researchers such as Chaitin (2001), it may be considered “a lack of structure” (Goldreich 2008, 285). Chaitin (2001, 18) defined randomness as something that “can't be compressed into a shorter description” and does not have a “concise theory that produces it”. More precisely, he stated that, in order to be considered random, something must be “algorithmically incompressible or irreducible” (Chaitin 2001, 111). The third approach, and the one Goldreich (2008, 286) focused on for his further discussion of pseudorandom number generation, views randomness as a concept that is “relative to the observer's abilities”. In short, this theory proposed that the notion of randomness resides within the perception itself.

Computational artistic systems often utilise these random number generators. The sequences produced by these generators will appear random, but are created using deterministic processes. Although generally referred to as random numbers, technically they are “pseudo-random” or “quasi-random”, with Knuth (1969, 3) further arguing that random numbers may only ever be apparently so. What we consider random numbers refers to natural phenomena and events whose processes of generation we cannot entirely explain. This may often be the reason why, rather than trying to further examine the processes themselves, we resort to using randomisers to simulate and mimic these processes.

In many cases, in particular when controlled and manipulated appropriately, this can lead to satisfactory results for the given purpose, but a certain degree of unpredictability remains. Although this unpredictability is something most generative artists will welcome, at least to a certain extent, it can also result in problems when applied to practices that have formerly been highly controlled and precise. In the case of video editing, the artist has to be very careful in order not to elicit the perception of randomness in the viewer. This would cause the viewer to disengage and thus render the process of generative video editing meaningless. Rather, the aim is to create conscious, as well as subconscious associations between the different images in a sequence by using randomness merely as a tool for the generation of variation.

3.3 Generative Sequencing of Narratives and Videos

The concept of generative music video composition discussed in this section is highly dependent on the availability of an appropriate software. Following Dewey (Hickman 1990), Brown (1999) suggests that using computational systems and software as tools for artistic practice actively influences the creation of artefacts. In this context, the primary motivations for the design of generative editing algorithms are (1) to extend video production practices and (2) to enhance the outcome of such practices.

Related research was undertaken by Nack (1996), who created *AUTEUR*, a prototypical system designed for the composition of humorous video sequences. In this work, Nack focused his approach on classic narrativity and the continuity system in editing, humour theory and artificial intelligence for autonomous decision-making. His framework of editing rules is derived from cinema and editing theory, as well as personal interviews with editors. Scenes were constructed primarily based on continuity editing and simple rules employing humorous elements categorised as slapstick comedy. These scenes were realised as non-verbal physical comedy, such as the banana-peel sketch. *AUTEUR* succeeded in creating credible humorous scenes based on a planner-based approach using hierarchical annotation systems for semantic, temporal and relational video content, as well as an ontological representation of narrative elements that establish appropriate links between video footage, and a number of strategies for narrative construction and editing. In addition, a simple mechanism for evaluating the humorous potential of a scene was used.

While Nack's system was based on reduced templates to adhere to conventions in continuity editing and the creation of non-verbal slapstick scenes, other projects explore algorithmic sequencing and non-linear presentation of narratives in a broader sense, albeit not necessarily in the domain of film and cinema, but, for instance, in multimedia storytelling.

Brooks (1999) developed a system named *Agent Stories* for algorithmic storytelling in order to explore and prove his concept of the metalinear narrative, which refers to a non-linear narrative presenting cinematic stories from multiple points of view. *Agent Stories* was designed as a tool for authors to construct their own metalinear narratives, which could subsequently be turned into multimedia applications. With the help of *Agent Stories*, an author would create a story abstract made up of a set of narrative primitives, i.e., the introduction of speakers and characters, conflict, negotiation, resolution, diversion and

ending (Brooks 1999, 101). This story abstract formed the basis for producing linear narratives from the metalinear narrative. The final stories are then assembled from story granules or clips by software agents employing behaviour-based artificial intelligence.

The research undertaken by Nack and Brooks highlights, among other things, the necessity of finding an efficient system of knowledge representation and semantic linkage, as well as a suitable algorithmic approach using various types of artificial intelligence in order to generate outcomes that adhere to the author's concept and are therefore coherent.

Other projects dealing with the sequencing of moving images have been realised in the fields of news and documentary editing, albeit focusing on information content and structuring instead of technical continuity (Ahanger 1999, Bocconi, Nack and Hardman 2005).

A different approach more strongly focusing on explorations in art than on the coherent structuring of information and strict adherence to editing convention as a design process, was chosen by Chambel et al. (2007) with their system *MovieGene*. The aim of this project was to find new creative possibilities for video editing by adopting evolutionary computation for the creation of video art. Chambel et al. created a system based on the combination of human judgement and computational evolution in order to achieve new and unexpected results. They emphasised that their method was not meant to replace human creativity, but rather to enhance it by presenting options an artist may not have thought of and letting him or her choose between different solutions.

Another example for the application of generative methods for the sequencing of moving images is ambient video art. Largely based on computational randomness, *Re:Cycle* by Bizzocchi (2011) is a system that creates ambient video sequences by assembling video footage using a variety of visual transitions. The aim of this project was to create ambient video applications that can run indefinitely, as well as to create variation. Bizzocchi also highlighted the compromise of giving up aesthetic control in order to gain access to the advantages offered by these technologies.

Clearly, adding generative methods to the video editing process opens up a wide field of both possibilities and potential issues.

Due to the precision required to create sequences that look like they were edited by a person, rather than a machine, it is questionable how effective the generation of video sequences that mimic, for instance, Hollywood cinema can be, since they must adhere to common conventions in continuity. Editing by a machine is often associated with a perception of randomness, as well as a lack of emotion and association. Although this direction in generative editing research remains an interesting field, I suggest that it may be more productive to apply generative editing methods, on the one hand, where the precision mentioned above may be less important than other functions in the video, e.g., in the case of news content, which naturally requires a greater emphasis on the type of information and its relative order in the video. On the other hand, it may also be beneficial to look at alternative and new types of video formats that naturally function within our new media environment and technological context, and can be built and established already with the application of generative editing methods in mind, rather than extending on well established modes of entertainment where this type of transformation may not necessarily be desired by viewers.

While some of music video's roots can be traced back to the beginning of the history of film, the music video we know today is a young and also aesthetically fleeting format. It changes with fashion, technology and any pop cultural influences that it can possibly draw on. However, what tends to remain is the discontinuous composition and generally experimental nature of music videos. Without the need for strict narratives and editing conventions, it is possible to approach the editing process in relatively new and unexpected ways.

4 Music Videos

"As a product of youth culture, the music video stakes all on its claim to coolness, and its embrace of stylistic extremity is a direct result."

(Austerlitz 2007, 3)

When music video first became a popular media format, it was said to be, e.g., "one of the most important emergent cultural forms in contemporary culture" with "a profound impact both on music, fashion and youth culture, and on the codes and forms that operate across television, film and advertising" (Frith, Goodwin and Grossberg 1993, np). In contrast to many other mainstream film and video formats where aesthetic conventions are much more standardised, music video has provided a variety of creative opportunities for directors, supported by the format's typical short duration of around four or five minutes, and a general lack of constraints.

As a media format, music video has been attributed a considerable impact on cinema and TV. In addition, it has also been suggested that it constitutes a prime representation of postmodernism in media because of its unpredictability, irony and destruction of convention (Frith, Goodwin and Grossberg 1993, Austerlitz 2007, Danesi 2008). For many young filmmakers growing up with MTV, music videos along with television commercials also allowed them to experiment, and to gain professional experience. Some of these filmmakers would later move on to cinema and other areas of television, thereby transferring some of their knowledge and experience into these areas, eventually leading to the creation of a new style, frequently referred to as MTV style, which Dancyger (2011) described as an editing style that is multi-layered and ambiguous rather than linear and continuous. Above all, MTV style is characterised by "the importance of creating a definite feeling state" (Dancyger 2011, 168). Considering a video format like music video, which is based on the effective creation of viewer engagement and an emotional state in a very short period of time, it only seems natural that some of the methods employed will eventually find their way into other areas of filmmaking.

This chapter examines the origin of music videos and how MTV style can be understood. As the medium used for experimentation in this thesis, music video editing, according to MTV style, is hereby highlighted to adhere to aesthetic dynamics distinct from other audio-visual forms.

Several developments promoted the emergence of music video as an independent video format. Officially, we can credit the invention of the music video format, as we know it today, to the launch of MTV (Music Television). However, although not known as music video at that time, the format evolved decades before MTV started broadcasting through various realisations, from “Phonoscènes”, to “Cinéphonies”, “Scopitones”, “Promos” and “Clips”, etc. (Schmitt 2010). Austerlitz (2007) suggested that music video has been one of various attempts to reunite music and vision, after their separation by radio and music recordings.

When sound was first introduced to film with the Kinetophone in the 1890s, many early examples depicted dances and marches, because these scenes provided a suitable means to both create visual rhythm and overcome issues concerning the close synchronisation of sound and image in these early apparatus. Altman (2004, 81) explained that, “whereas people making sounds requires exact instantaneous synchronisation, dancers and marching bands easily tolerate loose synchronization”.

An early example of a film that focused on visual rhythm rather than a narrative is Fernand Léger's *Ballet Mecanique* from 1924. Le Grice (1977) classified *Ballet Mecanique* as “systemic film”. In this context, Le Grice described the use of a system for visual composition as a replacement of a narrative, effectively resulting in a means of “eliciting a conscious structuring mode in the audience” (Le Grice 1977, 135). Similar techniques have found application in a variety of films. Alain Robbe-Grillet constructed narratives based on generative structures and musical rhythm (Armes 1981). He was interested in the exploration of causality, which, according to him “are really the same thing in a traditional narrative” (Robbe-Grillet and Morrissette 1977). A notable example of this line of work is the film *L'Eden et après* released in 1970.

We could say that music video's earliest ancestors are illustrated song films. Illustrated songs were a popular type of performance in vaudeville around the turn of the nineteenth century, with thematically associated images projected onto a screen behind a musician playing popular music. This popularity inspired film producers to imitate the concept and record films that were set to music (Altman 2004).

Although sound film slowly began to emerge as a new format at the beginning of the twentieth century, films were typically produced for exhibitions, vaudeville and nickelodeons

with musical accompaniment in mind. Apart from technological challenges, this may be because the images themselves had already been praised for their until then unknown realism, and film producers sought to merely accompany their impressive visuals, rather than to create an entirely realistic experience (Altman 2004). Film was still an emerging medium and it would take several decades until many of the conventions we are accustomed to in cinema today would finally be established.

During the 1940s until the mid 1970s, several jukebox machines playing music films became a relatively common source of entertainment in public places, most notably the Panoram and Scopitone (Keazor and Wübbena 2010). Films featured on these devices show a close relationship to the contemporary music video²⁹ due to the absence of a narrative-driven logic in favour of a performance-driven sequential progression. With their aesthetics and dynamics largely borrowed from the American musical, these films were also used to visualise popular music in new ways (Austerlitz 2007). Herzog (2007, 39) suggested that they should be regarded as “constellations of images” rather than linear sequences.

In contrast to MTV, the various types of jukeboxes introduced in this period never became very successful, and the featured films had been produced on very low budgets. The technology was never seen as much more than a fad (Herzog 2007). However, from the 1950s onwards, with the increasing popularisation of television, music films were also broadcasted on various TV stations, with one popular format being the Snader Telescription, which usually showed simple recordings of live performances (Austerlitz 2007). The genre was further influenced by music films such as The Beatles' *A Hard Day's Night* by Richard Lester. The concept was quickly imitated by contemporary artists, e.g., Pink Floyd, The Who and The Kinks, already affecting both TV and cinema productions with its ambiguous images and fast-paced editing (Dancyger 2011).

Although not properly categorised at that time, these precursors of music video slowly established basic aesthetic themes and altered viewers' awareness.

²⁹ Many examples of Scopitones can be found on YouTube:

Futurdada. “Nancy Sinatra These Boots are Made for Walking Scoptine”, *YouTube* video, 2:36. 7 November 2010. <https://www.youtube.com/watch?v=KgQKzFOidW4>.

historycomestolife. “Tell Him by The Exciters Scopitone in Stereo”, *YouTube* video, 2:30. 11 July 2010. <https://www.youtube.com/watch?v=io7taNUIlly4>.

ScopitoneDotCom. “Scopitone: Gale Garnett - “Small Potatoes” (S-1070)”, *YouTube* video, 3:15. 7 July 2013. https://www.youtube.com/watch?v=-wSW_bwJXvo.

4.1 Music Video on TV

In 1981, MTV launched as a twenty-four-hour-all-music television channel. MTV was meant to represent an alternative to radio, and to be a next step in the development of music-based media in order to support the music industry, which had reported a decrease in sales for several years (Wolfe 1983, Lynch 1984). Radio broadcasters in the U.S. had started to orient themselves towards audiences who were in their late twenties or older to provide a better basis for advertising (Straw 1993). However, this had created a gap for younger audiences. Therefore, MTV focused on a target audience of 12-34 year-olds by predominantly broadcasting new rock music. In the following years, the channel widened its scope, establishing numerous programs that featured different musical genres. A number of TV channels inspired by MTV emerged, and music video channels became a popular and successful worldwide phenomenon.

When contemplating on music video style, we typically think of fast edits and disjunctive shot juxtaposition. The music video has been compared to the TV commercial and described as being ephemeral, ambiguous and disjunctive (Frith, Goodwin and Grossberg 1993, Jhally 1995, Morse 1986, Vernallis 2004, Williams 2003). Goodwin (1992) argued that these characteristics were part of a deliberate design, not just to visually appeal to a specific generation of music video viewers, but also to provide a mode for repeated viewing, thereby preserving novelty. In contrast to Hollywood feature films, for instance, the viewer wasn't supposed to grasp everything going on in a music video at once. Rather, the central goal was to leave a stylistic impression. Vernallis (2004) stated that music video editing aims to preserve a sense of openness in order not to give priority to one particular element. Often the editing will be visible. Indeed, attention may be drawn towards the editing itself, which is against the basic ideal of continuity editing, a concept widely applied in narrative films.

Here, continuity refers to the spectator's visual perception and orientation concerning action, location, screen direction, scene, tone and dramaturgy. The continuity system is used to create a smooth flow over a series of shots in order to immerse the viewer in the narrative and diminish the attention drawn towards the technical aspects, e.g., production and editing (Bordwell and Thompson 2010, Dancyger 2011). Continuity is a central aspect in what Bordwell and Thompson (2010, 94) refer to as "classical Hollywood cinema", the dominant fictional filmmaking technique used, especially, in American studio films. Among the main

aspects of continuity in classical Hollywood cinema are chains of cause and effect which guide the narrative and are used to create a tight cinematic structure. Classical Hollywood cinema, according to Bordwell and Thompson (2010, 95), is also largely based on “objective story reality”, “unrestricted narration” and, finally, closure.

The continuity system and other conventions typical for classical Hollywood films are less dominant in music videos. Vernallis claimed this might be due to the possibility of the narrative dominating other elements in the video, which, according to her, is rarely the case. Rather, the music video is driven by its highly aestheticised nature, which foregrounds artistic and experimental imagery through framing and editing techniques, such as jump cuts and graphic match, which are rather uncommon in classical Hollywood cinema.

Many music videos can be described as progressing in a cyclical and episodic fashion, often following the song’s form (Vernallis 2004). While some music videos apply aspects of classic narratives by adopting a progressive structure with a clear beginning, middle and end, obvious chains of cause and effect comparable to Hollywood feature films remain a rare occurrence. Often, music videos will develop following what Vernallis (2004, 197) labelled a “slow disclosure”, with the purpose or consequences of actions gradually revealed in the video. Both Vernallis and Chion (1994), as well as Goodwin (1992), among others, referred to music video visuals and their progression as being of musical nature, not providing a stable foundation of visual linearity viewers may be accustomed to from traditional narratives. Instead, the images adopt an acoustic sound structure and become naturally “discontinuous and nonhomogenous” (McLuhan 2004, 71).

Chion (1994, 166) suggested that music video images are often viewed as “eye-assaulting” by “cinephiles” because of the “stroboscopic effect of the rapid editing”. Instead of providing a logical succession of events, the music video rather tends to illustrate a scene, feeling or even just a particular fashion or style. Often music videos will consist of both clear and ambiguous shot juxtapositions in their narrative structure as well as obvious and less obvious synchronisation points between music and visual. In order to better understand the video, the viewer may have to extrapolate and engage with the music video’s content more actively compared to a classic story-centred film.

Austerlitz (2010, 39), taking on a particular ironic perspective on some of music video’s typical characteristics, described music video as “[g]lorious trash, trash that occasionally

strives for the condition of art, but trash nonetheless". He further elaborated that music videos are "advertisements for music" with a "stated intent to move widgets with maximum efficiency". In this sense he described the nature of music video to be outlandish and rarely subtle.

A strong relation between music video and film or video sequences based on montage theory pioneered by the early Russian film directors, e.g., Eisenstein, Pudovkin and Kuleshov, has been suggested (Willemen 2002, Vernallis 2007). Vernallis, for instance, has drawn several connections between the two media formats, claiming that their common characteristics include a limitation of resources, which requires and evokes a certain kind of ingenuity (cp. Chapter 6.2.2).

The Russian filmmakers had to deal with film being a valuable material in addition to other constraints, such as the lack of film sound, which at that time had yet to be established, as well as a high degree of illiteracy among a majority of their audiences. Often these silent films were played to music as well. The production of meaning through images and their juxtaposition therefore received much attention.

Music video, in comparison, always constitutes a short format. Marketing requirements and the songs themselves may impose creative restrictions, in addition to the common need to showcase the artist(s), who will usually be seen performing the song.

Vernallis further suggested that in neither case these limiting factors had necessarily been treated as limitations by the filmmakers themselves. Instead, they just formed a framework allowing their innovation and creativity to unfold by using highly expressive visuals and exploring novel narrative structures.

In the late 1980s and early 1990s, a lot of research focused on the music video format, which had began to differentiate itself by abandoning conventions and establishing novel modes of narrative presentation. In the mid to late 1990s, however, music video had slowly started to disappear from television to make room for more lucrative programming. This development was quick in some countries, for example in the U.S., and slower in others, e.g., parts of Europe and Asia, but eventually the importance of television as a broadcasting medium has faded (Jones 2005).

4.2 Online Music Videos

After the music video had gradually disappeared from television, particularly in the early to mid 2000's, watching music videos online became increasingly prevalent. There are several reasons for this development. Initially, MTV had introduced new concepts for marketing to younger audiences and branding entire channels towards becoming a lifestyle choice, thereby transforming its status from a mere media outlet to an active source of popular culture.

Eventually, however, the music video turned out to be less profitable than other types of TV programming, as soon as more traditional television formats could be presented within the music video's marketing concepts.

Additionally, the Internet began to provide several advantages for watching music videos. Viewers could now access virtually any music video online, free of charge, any time they want. Most video platforms started to offer features such as a custom playlist creation and automated suggestions for similar videos a viewer might be interested in. Therefore, music video programming also became less desirable because viewers no longer had to settle for a preselected choice, but could make their own decisions instead (Jenkins 2006).

Following this development, contemporary music video research stresses the need to decontextualise music video and to examine how its dynamics and aesthetics have changed (Vernallis 2010, Edmond 2012). Vernallis, for instance, investigated the influence of the music video's new environment on the video format itself. She observed a continuation of music video's already inherent fragmentation that includes the phenomena that emerged with the appearance of viral online videos, such as intertextuality, parody and implications on graphic realisation. Edmond described the changing situation for music video with respect to its production and distribution, as well as its improved potential for various forms of advertisement and cross-promotion. Furthermore, Kern (2011) suggested that music video production should consider consumption on mobile devices and changing screen technologies.

Online media may have had an influence on music video, but many, if not all of its initial functions remain, such as the preservation of novelty to successfully promote music and capture viewers' interest for longer periods of time than just the music video's duration. An increasing number of online music videos specifically focus on novelty, i.e., the visualisation of an original idea that viewers will be more likely to share in order for the videos to go viral.

Following Jenkins³⁰ and Edmond (2012), we can quote OK Go's *Here It Goes Again*³¹ as a typical example. Edmond (2012, 306), in fact, described this video as "a watershed moment in recent music video history". The video shows the band performing a choreographed routine on a number of treadmills and consists of only one continuous shot to highlight its authenticity and to elicit a sense of surprise and amusement in the viewer (Edmond 2012). Increasingly, online videos have started to stress such unique features to attract attention. Viewers tend to be more likely to engage with this kind of content because of, in this case, its display of unusual skill and finesse, as well as by raising questions such as "How did they do it?" and "Did they really do it?". According to Jenkins,³² the "best YouTube content is content that is so unbelievable it has to be shared".

Edmond (2012) also used the example of *Here It Goes Again* as an argument against music video regarded as a postmodern phenomenon. In fact, she described this video as being "radically different" from postmodernist ideas because it is "coherent, focused, linear and intact" (Edmond 2012, 315). Furthermore, she argued that the video does not represent a new development, citing videos from the 1980s, such as *Girls on Film* by Duran Duran, *Material Girl* by Madonna and *Sledgehammer* by Peter Gabriel, as examples for a similar type of aesthetic and narrative progression.

This again demonstrates the wide range of different types of music videos, and that research often tends to focus on a few cases that exemplify certain patterns, rather than considering all manifestations. What, then, is the role of online video platforms such as YouTube in the current development of online music videos?

Online video, mostly hosted on YouTube, increasingly blurs the boundaries between different video genres and audio-visual characteristics associated with professional and non-professional production (Vernallis 2013). Online platforms such as YouTube provide an ever-growing body of videos that do not adhere to traditional filmmaking techniques and concepts

³⁰ Jenkins, Henry. "Youtube and the Vaudeville Aesthetic", *Confessions of an Aca-Fan - The Official Weblog of Henry Jenkins* (blog), 20 November 2006, http://henryjenkins.org/2006/11/youtube_and_the_vaudeville_aes.html.

³¹ emimusic. "OK Go - Here It Goes Again", YouTube video. 3:04. 26 February 2009. <https://www.youtube.com/watch?v=dTAAAsCNK7RA>

³² Jenkins, Henry. "Youtube and the Vaudeville Aesthetic", *Confessions of an Aca-Fan - The Official Weblog of Henry Jenkins* (blog), 20 November 2006, http://henryjenkins.org/2006/11/youtube_and_the_vaudeville_aes.html.

regarding continuity and seamlessness. Conventions for film and video production may therefore notably transform with evolving literacy in online media. Perhaps film and video started to diverge from conventional aesthetic systems for production and editing, just as these systems have finally become an established framework defining how moving images must be organised.

An essential aspect of YouTube videos and their success is the element of social engagement by sharing, discussing or even remixing and parodying YouTube videos. The success of these videos depends on the active engagement of viewers with their peers. This, however, does not constitute a new development. Jenkins has drawn a connection between online video and vaudeville concerning their general aesthetic, which is most of all based on variety and therefore known for affecting surprise and astonishment. Similarly, moving images have been considered a medium of innovation, social engagement and gathering in those times.

This may be a good opportunity to take a look back at (music) video's history, to risk an outlook on its future, and to realise the potential of moving images beyond established mainstream cinematic forms. While moving images on TV already transformed an essential part of popular culture and our daily lives, online media will further accelerate the development of these time-based media due to their perpetual availability and increasing importance for communication, entertainment and information.

4.3 The Future of Music Video

Vernallis (2013, 208) claimed that music video used to be more clearly definable in its earlier years as a “product of the record company in which images are put to a recorded pop song in order to sell the song”. But according to her, since music video has largely moved online, “none of this definition holds anymore”. Because virtually everyone can create and publish music videos online for any purpose or lack thereof, Vernallis redefined music video as “a relation of sound and image that we recognise as such”.

In the broad context of online media, moving-image formats and presentation outlets, this is an appropriate definition that takes many dimensions of media creation and influence on other digital media into account. In her definition, Vernallis does not differentiate between professional means of advertising and fan-made video mash-ups using arbitrary sound and video footage according to one's taste or otherwise creative concerns.

At this point, for the ongoing discussion within the scope of developing generative music video editing algorithms, we shall take a step back and consider music video in its more narrow and possibly more traditional sense as a promotional tool for music and musicians, and focus on a mostly professional and deliberately commercial production since the following discussions revolve around the potential development of software tools and techniques that facilitate the professional creation of generative music videos, and seek to imitate, transform and extend some of the established aesthetics in this creative practice. It should be noted that software solutions for consumers providing automated editing solutions are not considered at this stage. Within this project, the use of automation refers first and foremost to the addition of a further creative element that, in general, has not been part of the editing process before.

For designers, developers and artists alike, music video may offer interesting possibilities as an established means of experimentation in a short audio-visual format. Some authors already recognise music video as more than just a marketing tool. For many musicians it represents an additional layer of interpretation to a song, and an art form in itself (Wilscher 2011).

Without a doubt, music video will remain part of popular culture and digital media as long as popular music remains part of it as well. The important questions concerning music video address its further evolution and development. Austerlitz (2007) suggested that endeavours to reunite sound and image, which began to occur after the introduction of new technologies for recording both, may be governed by the desire for the *Gesamtkunstwerk*, an attempt to merge different art forms into one complete work, with each individual element elevated by its association to the other through the motivation of approaching “borderlessness” (Follett and Finger 2011, 3). Follett (2011, 87) argued that the idea of the *Gesamtkunstwerk* emerged due to a “deep original experience of alienation and disintegration” between “humanity and nature”. With technology, it seems, we do both: realising our human potential but also removing ourselves from our natural context, in this case, by fragmenting the objects we perceive into individual elements, i.e., sound, image, movement, progression. Is it possible that digital media help us to approach a state of borderlessness or a sense of completeness, which has not been offered by the media before, despite its ever increasing influence on our everyday lives?

Online media can be said to be inherently non-linear. In this context, Manovich (2001) referred to the database structure, which has started to take on much of the narrative's role and importance as a mode of organising information. Considering McLuhan's (2004) emphasis on the dominance of the visual sense in Western culture, we may be inclined to think that evolving technologies and media have opened up the way for our other senses to step forward, resulting in a more active engagement with different types of media and presentation.

Music video has established itself as a suitable means of experimentation and exploration in time-based online media. Some themes and methods that a new type of music videos, referred to as interactive music videos, have been exploring are, e.g., interactive narratives or choose-your-adventure narratives, explorable video spaces, videos that are, at least in part, generated by user input, and gamified videos based on crowdsourcing.

At the time of this writing, a number of companies has already emerged producing software tools facilitating the creation of interactive videos, e.g., Interlude and Rapt Media. It is therefore safe to say that non-linear and interactive videos will become more common in digital media. What remains to be seen is the types of interactivity and methods of sequence creation that will stand out from the rest and manage to engage users in the long term.

Viral mechanisms, which represent a modern version of word-of-mouth communication, are becoming more and more important for the distribution of online media, which are increasingly built around this premise and therefore, in a globalised world, also force media producers to find new ways for their videos to become more noticeable, more memorable and more attractive than other media artefacts. It is hard to say how this will continue to influence online media, but it can be assumed that more people will try to develop methods to increase the chances of content going viral. However, Jenkins, Ford and Green (2013) suggested that viral phenomena in online media may not be as predictable as some marketeers would like to claim and highlighted the importance of social and cultural context. We can only observe general dynamics among online media audiences, and they seem to increasingly resemble that of the vaudeville era.³³ Diversity, technological innovation and the unconventional are again

³³ Jenkins, Henry. "Youtube and the Vaudeville Aesthetic", *Confessions of an Aca-Fan - The Official Weblog of Henry Jenkins* (blog), 20 November 2006, http://henryjenkins.org/2006/11/youtube_and_the_vaudeville_aes.html.

taking centre stage. But this time, due to the new technological possibilities, it is the digital variety theatre where a worldwide audience is waiting to be entertained.

5 Online Music Video Analysis

“Music video is making a strong global comeback because of [YouTube … .] As the site’s number-one streamed content, music video consumption is dramatically up. It’s the perfect form to quickly set the pulse of our daily lives, as well as to grab a moment’s respite while websurfing and engaging in repetitive work. Music video clips on YouTube might help us gain the pulse of today’s world: perhaps in our heteroglot but connected environment, these clips will help global citizens discover a shared rhythm.”

(Vernallis 2013, 183)

In this section, two different methods of music video analysis will serve to validate and extend the theories concerning music video aesthetics and characteristics presented in Chapter 4. First, two sets of carefully selected music videos will be analysed in the form of individual case studies. Second, a larger body of music videos will be used to reveal overarching editing patterns and trends. Furthermore, the role of current media platforms and evolving technologies and their influence on these videos will be discussed.

Vernallis (2013) pointed out that (popular) music has become omnipresent in our lives, particularly with the introduction of mobile devices like iPods and smart phones, which provide constant soundtracks to our every action. Similarly, music videos, or some of their functions at least, have become increasingly ubiquitous in digital media, appearing in games, apps, and of course on social networking platforms and other websites. In addition, the production of music videos is no longer limited to professionals. Virtually anyone now has the ability to create and publish music videos, as mentioned in Chapter 2.

Whereas music video used to have a strong influence on TV and cinema, Vernallis suggested that online video platforms, in particular YouTube, are now influencing music video and other audio-visual formats in return. However, although MTV and music video are sometimes cited as the causes of film and cinema becoming increasingly disjunctive by abandoning classic continuity conventions and creating faster-paced sequences as well as using a larger number of close-up shots, Cutting et al. (2011) showed that the development in cinema has been following a gradual 75-year trend of increasing pace in editing and motion, as well as a decrease in shot luminance, since the 1950s. These trends, Cutting et al. suggested, can be regarded as linear trends in the evolution of media, film and video, and therefore do

not show a significant change after the introduction of MTV. They also traced this development back to the introduction of Hollywood-style sound film, since silent films displayed both similar editing patterns and similar average shot durations as films from the 1990s.

Therefore, while Hollywood film has seen a gradual evolution since it first developed into an established and recognisable form, the aesthetics of music video, e.g., its quicker pace and its common spectacle-driven progression, may not constitute a particularly unusual phenomenon in the evolution of moving image media, but rather represent a natural approach to the organisation of images. Cutting, DeLong and Nothelfer (2010) and Cutting et al. (2011) therefore argued that cinema is changing in these ways because filmmakers are gaining better knowledge on the underlying mechanisms to influence viewer engagement, and are therefore better able to control viewers' attention.³⁴

In this context, we also have to consider the influence of technology on media. In Chapter 2 and Chapter 3, it was shown that technology and tools are essential aspects of human nature, and that technology, at the very least, will necessarily influence artistically orientated works. Monaco (1981, 49) suggested that technology and media aesthetics form an active relationship, where aesthetic preferences change according to new technology developments, and sometimes technologies have to change in order to facilitate new aesthetic requirements (cp. Cavell 1979, Carroll 1996 and Rodowick 2007). According to Monaco, film's codes and tropes, in contrast to other forms of "art", strongly depend on the medium's technology.

Edmond (2012) argued that, despite the varying transformations of media through technology, music video aesthetics may not have changed that much at all, and Vernalis (2013) added that many of YouTube's typical characteristics had first been established with television, whereby music video has had a strong influence. In addition, it is important to consider YouTube's technical background, restrictions, and mode of operation, including the videos' brevity as well as the social interaction, which is vital to their distribution. Vernalis (2013, 14) also observed that all media "from post-classical cinema to music video and commercials start to resemble or refer to YouTube".

34 Cp. Miller, Greg. "Data from a Century of Cinema Reveals How Movies Have Evolved", *Wired*, 9 August 2014, <http://www.wired.com/2014/09/cinema-is-evolving/>.

According to Vernalis' (2013), temporality is the main aspect that is changing in our media landscape. Other aspects that are changing include display formats, e.g., technical aspects of quality regarding screen size and compression, as well as social engagement surrounding the media.

Time behaves differently on the Internet. Software-based media allow for non-linear consumption on the one hand and fast access of an unprecedented amount of information on the other hand (Manovich 2013). Videos are becoming shorter, and a stronger sense of rhythm seems to arise with the possibility of switching from one clip to the next instantaneously. Users can exercise control over the video's progression by intuitively skipping within their timelines, or opening and closing parallel browser tabs to engage in all kind of content at the same time (Vernalis 2013).

To ensure a high download speed, the uninterrupted streaming and an ideal visibility on mobile screens, aesthetic preferences for online video prioritise the use of close-ups, minimal movement and two-dimensionality (Vernalis 2013). Whereas Edmond (2012) stated that music video has always featured both continuous and discontinuous structures, and that the aesthetics of online music videos may not be particularly different compared to music videos predating the Internet, Vernalis (2013) observed an increase in discontinuity, simultaneity and causal ambiguity.

This again demonstrates the difficulty to define editing preferences for music video. In many ways, music video editing still adheres to the initial design from its inception by MTV in the early 1980s; in other ways it seems to have developed a life of its own, with technological experimentation allowing it to explore new dimensions, while also establishing its dominance in online media. Later in this chapter, I will propose a framework for such a description based on a detailed music video analysis.

Inspired by OK Go's music video for *Here It Goes Again*,³⁵ Jenkins even compared online video to the popular twentieth century vaudeville variety shows. Vaudeville was a popular form of stage entertainment and provided its audience with a number of vastly different performances and demonstrations featuring musicians, comedians, lectures and movies, etc. Faced with this kind of direct competition, each performer would strive to be the most

³⁵ emimusic. "OK Go - Here It Goes Again", YouTube video, 3:04, 26 February 2009, <https://www.youtube.com/watch?v=dTAAsCNK7RA>.

memorable. It is easy to observe a similar behaviour in online media, where each artefact now has to compete with the rest of the world, rather than an enclosed circle of contenders. These media artefacts need to be found and shared according to “processes of social appraisal” (Jenkins, Ford and Green 2013, 195). Viral marketing and the potential “spreadability” of an artefact are now important aspects in the creation of media. After gaining enormous popularity online, in some cases, artefacts may even transcend their own representation and turn into a generalised imitable idea – a concept now widely referred to as Internet memes. The term “meme” was initially introduced by Richard Dawkins and later used to denote these online phenomena because their function is similar to the function of Dawkins’ memes, who defined them as cultural “replicators” or “units of imitation” (Dawkins 1976, 192, Shifman 2014). Memes can therefore accelerate the spread of an artefact by turning it into an idea and a concept that allows audiences to participate.

Before examining music videos in more detail, the next section will briefly discuss structure and narrative with a particular focus on music video to support the following case studies and analyses.

5.1 Structure and Narrative

In Chapter 4, music video aesthetics were described as ephemeral, ambiguous and disjunctive. The form’s structural progression was characterised as cyclical and episodic, whereby obvious narratives with clear chains of cause and effect remain a rare occurrence. Thus, many of the descriptions given so far rather demonstrated what music video is not: clear, logical and easy to understand. This section therefore attempts to provide a more detailed description of what music video *is* by referring to its narrative or non-narrative structures and context.

While some authors view music video as a form of art or an artistic extension of musical practice, first and foremost, music videos are created to serve as a promotional tool that frames the musician and creates an image in order to sell the music (Strachan 2006, Wilscher 2011). Most important for the success of music videos of many high-selling acts is the creation of an almost mythical character portrayed by the musician, an object of admiration as well as adoration (Peverini 2010). Narrative clarity is therefore only a secondary priority. Instead, music video editing is often used to “underscore non-narrative visual structure and form such structures on its own” (Vernallis 2001, 21). Vernallis (2004, 53) further elaborated that, while music videos may appear discontinuous at first glance, the types of structures established in

them usually interlace a number of narrative and non-narrative strands in order to purposefully create a densely organised and largely unpredictable sequence. Rather than following a traditional narrative, music videos tend to be “processually structured” in order to underline and support the music (Vernallis 2013, 97).

Jones (1988) referred to music video as being part of a new narrative form primarily based on sight, denoting this form the ‘digital narrative’ defined by a non-linear structure, an increased openness to interpretation and a presentational form which delivers information in associative bits. This description is similar to Manovich’s (2001) definition of the database form in *The Language of New Media*. Manovich suggested that both novel and cinema have shaped cultural expression as a whole. In this sense, computers and software again are influencing our perception and creation of artefacts that operate via interfaces and which allow users to “perform various operations – view, navigate, search” (Manovich 2001, 219).

The following sections attempt to examine music video aesthetics by adopting the idea of a new narrative structure and with reference to film narratology in order to further differentiate music video from cinema.

Definitions of what a narrative is and recounts of elements that are deemed necessary for something to constitute a narrative may vary greatly. Stam, Burgoyne and Flitterman-Lewis (1992, 69) suggested a definition of the term ‘narrative’ by referring to theorists such as Todorov (1971), Rimmon-Kenan (1989), and Prince ([1987] 2003): “Narrative can be understood as the recounting of two or more events … that are logically connected, occur over time, and are linked by a consistent subject into a whole”.

Todorov (1971), who introduced the term narratology as a descriptor for narrative analysis, proposed two basic narrative principles: a story is distinguished on the one hand by the succession of narrative elements and on the other by their transformation. According to Todorov, narratives have a beginning, middle and end, a structure that is typically realised through the following five stages: (1) a state of equilibrium, (2) a disruption of this state, (3) a recognition of this disruption, (4) an attempt to undo or repair the disruption, and (5) the re-establishment of equilibrium.

Based on earlier developments in semiotics, structuralism and Russian Formalism, narratology largely derives from literature analysis but was also applied to early film theory. In

the 1920s, theorists like Victor Shklovsky and Juri Tynianov significantly influenced the narratological discussion of film, which sought to provide a better understanding of narrative logic, time and space in the context of the Formalist narratological terms revolving around fabula and syuzhet (Stam, Burgoyne and Flitterman-Lewis 1992).

In the 1980s, these ideas were once again adopted by Bordwell (1985) in *The Narration in the Fiction Film*. Bordwell's work highlighted some difficulties that arise when using strict narratological frameworks as those briefly outlined above. Music video and other television formats tend to lack these narrative structures and can therefore not be addressed in the same ways as fiction film. Similarly to Jones' definition of the digital narrative, Nelson (1997), for instance, established the idea of the flexi-narrative when referring to television drama shows. Therefore, while narratological theories can help us to understand the traditional narrative, we may need to rely on different and more open definitions when studying music video.

Abbott (2008, 3) described narrative as “the principal way in which our species organises its understanding of time”, considering narrative a cognitive mechanism that activates, for instance, even when only looking at a photo. Instead of just seeing a depiction of perhaps landscapes, buildings, rooms and various figures, our minds may automatically associate possible events and actions that have led to this representation of a single moment in time. In this context, we can conclude that our experience and knowledge tend to fill in the gaps left behind by what we can't see and don't know. Abbott (2008, 13) hereby followed a rather general approach by defining ‘narrative’ as “the representation of an event or a series of events”. An event, in turn, constitutes an action unfolding in time.

Branigan (1992, 19) investigated several types of narrative and suggested that “although the essence of narrative is a presentation of systematic change through a cause and effect teleology”, it is also possible to consider other ways of organising information as part of the narrative. Branigan outlined six ways of association, from non-organised data to more complex forms of narrative:

- (1) Heaps - A collection of practically random information assembled by chance.
- (2) Catalogues - A collection of information according to a central theme.
- (3) Episodes - A collection of information that form consequences according to a central situation.
- (4) Unfocused chains - A series of cause and effects without a central theme.

- (5) Focused chains – A series of cause and effects with a central theme.
- (6) Simple narratives – A collection of episodes in a focused chain.

Branigan (1992, 20) pointed out that these ways of associating information make use of different “types of time”, either by presenting events in an atemporal and simultaneous fashion, or by placing them in a more consequential structure. He concluded that “simultaneity, duration, causality, and order are not simply four items in a taxonomy of time, but are the results of four specific ways of processing data”.

By viewing classic narrative as only one type of temporal organisation, a wider range of possibility opens up for the creation of video sequences, which inherently have to unfold in time. Cubitt (2004, 38) elaborated that narrative “is not an essential quality of film, but only a potential and secondary quality arising from the production of time in the differentiation within and between frames”. In Chapter 4, as well as at the beginning of this chapter, it was suggested that online media may in many ways be compared to theatrical entertainment such as vaudeville. In this sense, Cubitt described film’s early aesthetics as eliciting a sense of amazement and wonder, with viewers seeing themselves as participants in an event that must be shared and communicated. The narrative, therefore, had not been a primary concern in the early days of cinema.

Similarly to Jones (1988), Dancyger (2011) asserted that a non-linear narrative is going to become more dominant in contemporary film and video.

Chapter 4 differentiated music video aesthetics from Bordwell and Thompson’s (2010) definition of classical Hollywood cinema. The comparison was mostly based on the application of continuity, which is a common means, especially in mainstream cinematic forms, to control the viewer’s attention and orient him or her toward the depicted narrative. The discussion above asserted that music video progression will usually focus on the realisation of a visual rhythm that dynamically connects and disconnects the images, their form and tempo according to motion and progression with the song’s musical structure (Rabiger 1989, Vernallis 2001, Vernallis 2004). Following these assertions, the investigation in this chapter focuses on music video’s evolving aesthetic in online media. Sibilla (2010) suggested that music videos put less emphasis on visual quality, but on highlighting an original idea. With a particular focus on YouTube, Vernallis (2013, 14) attributed changes in

music video aesthetics to a move towards “Youtube-ification” in media, which is predominantly characterised by a “do-it-yourself look or aesthetic”.

It is hard to confirm whether the conventions and common practices of music video conception and creation have undergone a significant transformation since digital media has become its predominant source of distribution. While many of music video’s aesthetics prevail as its very design seems to suit the nature of online media and video, the literature suggests the possibility that music video, as part of digital media, is becoming more discontinuous, fragmented and self-reflexive (Keazor & Wübbena 2010, Edmond 2012, Vernallis 2013).

In the previous chapter, the term ‘MTV style’ has been introduced and used to describe music video aesthetics that seemed to influence cinema and television formats, other than the music video itself. But according to Cutting et al. (2011), several claims made towards MTV style may have to be reinterpreted. As mentioned above, Cutting et al. (2011) have shown that several parameters of film aesthetics, e.g., the overall luminance of shots, editing pace and rhythm have undergone a linear development in the last 75 years. The introduction of MTV has not caused spikes in any of these parameters, suggesting that aesthetics associated with both MTV and YouTube may instead constitute natural developments in audio-visual media. The research conducted by Cutting et al. (2011) demonstrated that these media and their various forms are as young as they are culturally influential. Therefore they are undergoing a comparatively rapid change and evolution due to artistic intentions on the one hand, and optimised viewer engagement on the other.

Music video aesthetics have primarily been based on non-cinematic forms of moving images, such as experimental film and video art, as well as television commercials (Le Grice 1977, Wardrip-Fruin and Montfort 2003). As a result of these influences, music video has developed its own system of signs and tropes. Referring to music video as a “strictly codified form”, Peverini (2010, 148) drew attention to the importance of the performer, specifically the representation of body and icon, as the primary means to “install and to illustrate an adoration at the same time” (Peverini 2010, 143). Jhally (1995, 84) supported this idea from an advertising perspective. To him, music videos are clearly utilising “techniques learned from the marketing of products” by creating videos that are designed to gain attention and are “visually pleasurable” for audiences. Rather than foregrounding the product itself, advertising, according to Jhally, is mostly based on the depiction of people who “stand for reigning social

values” and on conveying emotion (Jhally 1995, 78). In the case of music video, this concept is based on showcasing the musician and placing him or her in a pop-cultural and ideological context that audiences want to identify with, thereby turning the music into a lifestyle element.

The musician will therefore usually be placed in abstract and symbolic locations that provide the basis for establishing an emotional state. Vague narratives and stylised performances are often highlighted through the frequent use of close-ups, unless important actions in a communal context are shown. Many music videos will seek to find a mode of communication with the audiences by demonstrating imitable actions (often simple dance routines) and via direct camera interactions with the camera during performance elements in the video. Often music videos will feature both, some form of a narrative as well as performance elements. This might be helpful to embed a strong association between the emotion conveyed in the video and the performing musician. Similarly, using fluctuations in tight and loose synchronisation between song and video, both can be foregrounded alternately or together to guide the viewers’ attention (Dancyger 2011, Vernallis 2013).

While these patterns and codes remain active for linear music video, the widespread use of online media has also introduced new concepts that now influence media production. In particular, spreadability and viral marketing will now permeate any discussion of online content and influence its creation (Jenkins, Ford and Green 2013, Shifman 2014).

With regard to these concepts, two basic approaches to optimise music video marketing based on maximising exposure and spreadability emerged: (1) the use of highly expressive, humorous, or otherwise unusual visual content, and (2) the implementation of interactive or non-linear methods in order to provide a different and potentially more immersive viewing experience.

Jenkins, Ford and Green (2013) pointed out the unpredictable nature of the processes surrounding the spreading of content. An increasing number of online advertising campaigns and commercials are designed towards this goal, and in recent years a growing number of researchers began to study the mechanisms and social processes behind viral online content (Berger and Milkman 2012, John 2012, Shifman 2014, Wiggins and Bowers 2014).

For music video, many cases remain where the song itself is the primary reason for its spreading. In this case, music video takes on the more passive role of completing the song's online presence. In fact, many of the most watched music videos on YouTube do not appear to be different at all compared to music videos created solely for television several years earlier.³⁶

The following sections will feature both linear and interactive music videos.

Chapter 5.2 will examine three linear music videos, which have been released between 2010 and 2012, and were therefore primarily distributed online. These three videos were selected because all of them achieved commercial success, but received popularity through different means.

However, a change in the above-mentioned relationships between song, video and audience can be observed for interactive or non-linear music videos. These "videos" introduce new modes of viewing or music video consumption, new aesthetics, progressive dynamics and a stronger focus on the visual rather than the music. Chapter 5.3 therefore discusses three different interactive and, to some extent, non-linear music videos produced between 2010 and 2014. These videos were chosen because of the different methods employed during their production which utilise the potential of current technological developments, and the creative ideas in this context.

5.2 Linear Online Music Videos

The three videos discussed in this section are *Gangnam Style* by Psy, *Big Jet Plane* by Angus & Julia Stone, and *Call Me Maybe* by Carly Rae Jepsen. These videos were chosen because all of them achieved at least some level of commercial success. Both *Gangnam Style* and *Call Me Maybe* represent successful viral videos that turned into memes. While *Gangnam Style*'s popularity was largely ascribed to its outlandish video, *Call Me Maybe* turned into a viral success because of the song's memorable lyrics and an effective promotion based on social media. Both became memes in 2013. However, whereas *Gangnam Style* triggered imitations that largely referred to the music video, the *Call Me Maybe* meme spread in two distinct

36 MyTop100Videos. "Most Viewed Videos of All Time • (Over 100 million views)", *YouTube*, https://www.youtube.com/playlist?list=PLirAqAtl_h2r5g8xGajEwdXd3x1sZh8hC.

directions: (1) covers and lip-dubs of the song, and (2) jokes and images revolving around the song's chorus (cp. Figure 5.2.1).

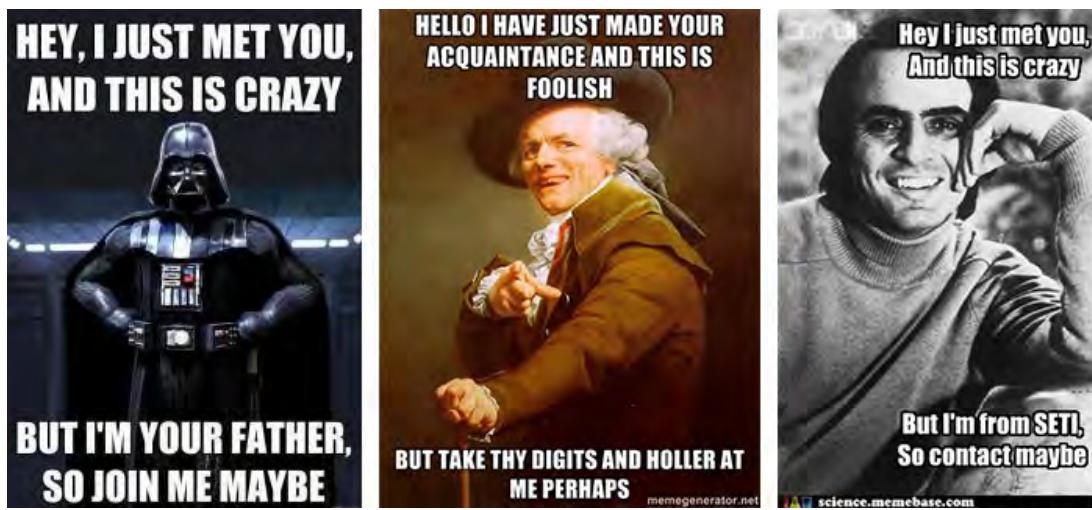


Figure 5.2.1: Examples for snowclone captions for *Call Me Maybe* (Know Your Meme. *Call Me Maybe*, <http://knowyourmeme.com/memes/call-me-maybe>).

Big Jet Plane was a chart success in Australia, but neither a worldwide, nor a viral hit. The music video for this song was included to provide a less popular example of a different musical genre and target audience, and which therefore may adhere to different visual aesthetics. The video for *Big Jet Plane* was also included in the analysis presented in Chapter 5.3, and used for first experiments, as outlined in Chapter 6.3.

Since the introduction of Flash video in 2002, which has provided increasingly feasible processes for compression and streaming of video material, a number of online video platforms have emerged. However, YouTube was and has been the most successful platform since its introduction in 2005. At the time of this writing, YouTube ranks as the third most popular website³⁷ after Google Search and Facebook. According to official YouTube statistics,³⁸ the site logs over six billion hours of viewed video with more than one billion unique users per month from countries all over the world.

Increasing interest has been directed towards YouTube for its role both in digital media and as a central element in popular culture. It has been described as both archive and video library, and some discussions revolve around its potential of replacing television's status in

37 Alexa. "youtube.com", Alexa, <http://www.alexa.com/siteinfo/youtube.com>.

38 YouTube. "Statistics", YouTube, <https://www.youtube.com/yt/press/statistics.html>.

popular culture (Snickars and Vonderau 2009, Uricchio 2009). Jenkins³⁹ suggested that YouTube is, above all and despite the increasing cooperation with commercial broadcasters, a grassroots development. In this sense, Jenkins elaborated that YouTube's instantaneous success derives from cultural dynamics and society's readiness for an online video outlet such as YouTube. Increasingly, the platform may act as a global conglomerate that takes on future roles of television, as much as that of a video archive and vlogger platform, among other functions.

YouTube lives off its users' active engagement and associated viral mechanisms. An essential characteristic of YouTube videos that makes this possible is their brevity and openness to all authors and audiences. Many videos can be quickly watched between other activities and are therefore often shared as short sources of entertainment.

Extending on definitions by Berger and Milkman (2012), Shifman (2014, 66) identified six factors that increase the chances of social interaction around online media content:

- 1) positivity,
- 2) provocation of high-arousal emotions,
- 3) participation,
- 4) packaging,
- 5) prestige, and
- 6) positioning.

In this sense, users share content that might be interesting to their peers or is likely to improve their mood, but also to present themselves in a particular way. Of similar importance are the content creators and those who are sharing the content. Articles written by famous celebrities or content shared by such celebrities via social media is more likely to receive a larger audience and convince users of the content's quality.

Shifman (2014, 74) further isolated several common features found in memetic videos:

39 Jenkins, Henry. "How YouTube Became OurTube", *Confessions of an Aca-Fan, The Official Weblog of Henry Jenkins*, 18 October 2010,
http://henryjenkins.org/2010/10/how_youtube_became_ourtube.html.

- 1) a focus on ordinary people,
- 2) flawed masculinity,
- 3) humour,
- 4) simplicity,
- 5) repetitiveness, and
- 6) whimsical content.

Shifman explained these features based on the viral mechanisms surrounding the *Gangnam Style* music video, which he referred to as an “ultramemetic video”. According to him, *Gangnam Style* combines all of the above features.

Vernallis (2013) claimed that music video is YouTube’s most viewed content, and indeed, further discussion in this section will confirm this argument. Vernallis attributed the popularity of music videos on YouTube to a general change that has been taking place in our everyday lives, stimulated by mass media. Declaring that “much media, across platforms and genres, driven by close audio-visual relations, are not what we grew up with”, Vernallis (2013, 6) observed an increasing musicality in moving images, mentioning the 2008 Obama campaign clip *Yes We Can*⁴⁰ as a notable example. Vernallis further identified eight features that characterise YouTube videos:

- 1) pulse, reiteration, and other forms of musicality,
- 2) irreality and weightlessness,
- 3) scale and graphic values,
- 4) unusual causal relations,
- 5) variability and intertextuality,
- 6) humour and parody,
- 7) volubility and other condensation, and
- 8) formal replication of the web.

Compared to the last few decades of moving images on television, these characteristics imply a change of temporality in video content that turns viewing into a rhythmic reiterative cycle of short image sequences. This phenomenon is not isolated to YouTube, but applies to online video in general. For instance, the social networking platform Vine has pushed the

40 WeCan08. “Yes We Can - Barack Obama Music Video”, *YouTube* video, 4:30 2 February 2008, <https://www.youtube.com/watch?v=jjXyqcx-mYY>.

development of online video in conjunction with social media. The platform introduced the six-second video which, once loaded in the application or on a website, is played repeatedly until the user navigates elsewhere. Repetition, reiteration and their inherent rhythm become a central aesthetic element of these videos. We can assume that various online video formats will continue to influence each other, depending on their success, i.e., amount of user response. Currently, social engagement and communication seem to be the main driving forces of the developments in this direction.

As mentioned above, music and therefore music video have played an important role on YouTube (Edmond 2012, Mjos 2012, Vernallis 2013). At the time of this writing, out of the list of the 100 most watched videos on YouTube,⁴¹ an auto-updated list based on YouTube's viewer count, 96 can be classified as popular music videos, the other four can be classified as home video footage (*Charlie bit my finger - again*⁴²), film and animation (*Masha and The Bear (Episode 17)*⁴³), children's music (*Twinkle twinkle little star*⁴⁴) and other recorded footage (*Evolution of Dance*⁴⁵). In addition, the most popular YouTube channel is the Music channel with over 88 million subscribers⁴⁶ at the time of this writing.

YouTube's first music awards show was held in 2013 in New York, once again demonstrating the importance of music videos to YouTube and underscoring the significance of social engagement, since all winners were determined by YouTube's users via online polls rather than a selected jury.⁴⁷ The show was directed by Spike Jonze who intended it to be unscripted

41 MyTop100Videos. "Most Viewed Videos of All Time • (Over 100 million views)", *YouTube*, https://www.youtube.com/playlist?list=PLirAqAtl_h2r5g8xGajEwdXd3x1sZh8hC.

42 HDCYT. "Charlie bit my finger - again !", *YouTube* video, 0:56 22 May 2007, https://www.youtube.com/watch?v=_OBlgSz8sSM.

43 Get Movies. "Маша и Медведь - Маша плюс каша (Серия 17) | Masha and The Bear (Episode 17)", *YouTube* video, 7:09, 31 January 2012, <https://www.youtube.com/watch?v=KYniUCGPGls>.

44 Super Simple Songs. "Twinkle Twinkle Little Star", *YouTube* video, 2:33, 5 September 2010, <https://www.youtube.com/watch?v=yCjJyiqpAuU>.

45 Judson Laippy. "Evolution of Dance", *YouTube* video, 6:00, 6 April 2014, <https://www.youtube.com/watch?v=dMH0bHeiRNg>.

46 MyTop100Videos. "Popular Channels", *YouTube*, <https://www.youtube.com/user/MyTop10Videos/channels>.

47 Spangler, Todd. "YouTube Music Awards Nominees Announced", *Variety*, 21 October, 2013, <http://variety.com/2013/digital/news/youtube-music-awards-nominees-announced-1200749352/>

Hernandez, Brian Anthony. "YouTube Music Awards Reveals Nominees and Opens Social Voting", *Mashable*, 23 October, 2013, <http://mashable.com/2013/10/22/youtube-music-awards-nominees-voting/>

and spontaneous in order to reflect YouTube's general nature and aesthetics as best as possible.

While several scholars predicted an uncertain future for music video by the early 2000s, we can safely assert that it has found a new home in online media for some time to come (Vernallis 2004, Keazor & Wübbena 2010, Vernallis 2013).

Extending on previous descriptions and analyses, the following case studies will therefore provide a glimpse on music video aesthetics in the era of online video and YouTube. While these case studies are far from exhaustive, they can help us to better understand music video progression, patterns, and overall aesthetics and prepare an understanding of some of music video's essential characteristics within the context of the following research.

In her book *Experiencing Music Video: Aesthetics and Cultural Context* (2004) Carol Vernallis provided some of the most in-depth analyses and case studies of music video aesthetics, progression and editing patterns by investigating videos such as *Cherish* by Madonna, *Get Off* by Prince and *Mercy St.* by Peter Gabriel. Vernallis examined how aspects like flow and continuity are realised in these music videos from an editing perspective. With regard to audio-visual connection and relationship, Vernallis particularly focused on the musical form and explained the video's progression, points of synchronisation and disconnection. Finally, she also discussed the role of narrative, space and the performer, as well as self-reflexivity and underlying values such as race, gender and sexuality.

The case studies presented in this section will predominantly focus on content structure and editing techniques. They will examine how music video's aesthetics and their promotional role may have changed as part of the online media environment in conjunction with social networking.

Similar to the music video analyses conducted by Björnberg (1994) and Vernallis (2004), the following case studies will strongly focus on the sectional song structures. Both Björnberg and Vernallis argued for a stronger focus on musicality and audio-visual connection in music video analysis based on the assertion that, since the music always precedes video production, the visual sequence necessarily follows or complements the songs. This constitutes an interesting perspective in the context of this research, since gaining a better understanding of structural definitions of music video and how these definitions can be translated and

communicated to an algorithmic system has been the primary reason for these analyses. As a basis for the practical experiments and the development of generative editing concepts as well as software development, several technical parameters received special attention throughout this research to guide formal definitions in generative music video editing. Audio-visual synchronisation, with regard to lip-synching and audio-visual rhythm, is an essential area of inquiry for the development of algorithmic editing methods, and is therefore treated as a key aspect in the following analyses. Other parameters that are focused on include the overall editing pace and changes in visual content throughout a sequence. These parameters influence the development of algorithms with regard to the amount of screenable footage that has to be produced for the generative video, as well as the placement and frequency of algorithmic decisions and structuring during the editing process. Furthermore, the investigation aimed to identify overarching structural patterns which may help to describe individual functions for certain features required during generative music video editing.

The case studies will show that many of music video's essential aesthetics outlined by Vernallis in 2004 remain prevalent, if not dominant. The following three examples share an orientation to the song's form by using clear changes in visual characteristics. Although this method is not applied consistently throughout the videos, it was, without exception, used to highlight the songs' bridge sections, which will usually constitute the most dramatic change in the song. In some cases, the changes in visual characteristics on sectional transitions are less dramatic than in their temporal juxtaposition. In addition to changes in content, the editing will tend to apply temporal highlights by increasing or decreasing the relative amount of cuts on sectional transitions. Since all of the following three videos feature some level of narrative progression, the bridge sections also introduce both a change in progression and the narrative's resolution or at least a change in direction that may transform the emotional response and elicit feelings of conclusion.

Another assertion by Vernallis (2004, 195) that has been confirmed in the analyses of the following music videos is that music videos might display closer audio-visual synchronisation towards the beginning and become "more tenuous" in the final section. While some points of clear synchronisation remain, they tend to appear less frequently in preparation for the music video's conclusion, which in this sense will often be intensified by showing increasingly expressive images, or by repeating and combining different scenes shown previously in the video.

The following process of focusing on musical form and sectional divisions has proven to be useful for isolating editing patterns and understanding changes in narrative and dynamic editing progression.

5.2.1 Gangnam Style

In 2012, *Gangnam Style*⁴⁸ (Figure 5.2.2) was the most watched as well as most liked video on YouTube.⁴⁹ TIME and Rolling Stone called it the best music video of 2012.⁵⁰ It became the first video in Internet history to be viewed more than one billion times and, among other descriptions, was referred to as the “greatest cultural masterpiece of 2012”.⁵¹ Most recently, in December 2014, the video exceeded 2 billion views,⁵² a number so large that YouTube was forced to upgrade its view counter from a 32 to a 64-bit integer.

48 officialpsy. “PSY - GANGNAM STYLE (강남스타일) M/V”, *YouTube* video, 4:12, 15 July 2012, <https://www.youtube.com/watch?v=9bZkp7q19f0>.

49 Lipshutz, Jason. “PSY’s ‘Gangnam Style’ Passes Justin Bieber’s ‘Baby’ For YouTube Crown”, *Billboard*, 24 November 2012, <http://www.billboard.com/articles/columns/k-town/474032/psy-gangnam-style-passes-justin-bieber-baby-for-youtube-crown>.

Barrett, Dan. “‘Gangnam Style’ Holds Guinness World Record for most ‘liked’ video in YouTube history”, *Guinness World Records*, 20 September 2012, <http://www.guinnessworldrecords.com/news/2012/9/gangnam-style-now-most-liked-video-in-youtube-history-44977/>.

50 Locker, Melissa. “9 Best Music Videos of 2012”, *TIME*, 17 December 2012, <http://entertainment.time.com/2012/12/18/9-best-music-videos-of-2012/slide/psy-gangnam-style/>.

Rolling Stone. “Best Music Videos of 2012”, *Rolling Stone*, 7 December 2012, <http://www.rollingstone.com/music/pictures/best-videos-of-2012-20121207>.

51 Guinness World Record News. “PSY’s Gangnam Style becomes first video to be viewed 1 billion times on YouTube”, *Guinness World Records*, 20 December 2012, <http://www.guinnessworldrecords.com/news/2012/12/psy-gangnam-style-becomes-first-video-to-be-viewed-1-billion-times-on-youtube-46462/>.

Johnson, Boris. “How Gangnam Style and Fifty Shades gave culture a spanking”, *The Telegraph*, 23 December 2012, <http://www.telegraph.co.uk/comment/columnists/borisjohnson/9764197/How-Gangnam-Style-and-Fifty-Shades-gave-culture-a-spanking.html>.

52 YouTube. *Google Plus* Post, 2 December 2014, <https://plus.google.com/+youtube/posts/BUXfdWqu86Q>.



Figure 5.2.2: Screenshots of the Gangnam Style music video (officialPsy, "PSY - GANGNAM STYLE (강남스타일) M/V, <https://www.youtube.com/watch?v=9bZkp7q19f0>).

The video's success has been ascribed to the catchy music and engaging video, characterised by colourful visuals and the trademark “invisible horse dance”, which is said to be easily imitable and succeeding in making a peculiar dance move look “cool”.⁵³ This was confirmed by an analysis of six threads of online discussion boards with the topic “Why is Gangnam Style so popular?”⁵⁴ Three of the most frequent answers were the song and video being “fun” to watch or “funny”, the song being catchy, and that these aspects were further supported by the video’s signature dance. Vernallis (2013, 143) observed that the video creates a sense of familiarity and humour by parodying earlier music video tropes. According to her, in addition to its humorous representation and easy method of physical participation with the invisible horse dance, it also creates a sense of community and associative location.

The video shows the musician Psy as the protagonist spending his time in the Gangnam district of Seoul, supposedly looking for a beautiful classy woman. The video is highly ironic, since Psy is shown to be the opposite of what is considered to be “Gangnam style”, a term that

53 Judkis, Maura. “‘Gangnam Style’s’ invisible horse dance craze: Giddyup”, *The Washington Post*, 23 August 2012, http://www.washingtonpost.com/blogs/arts-post/post/gangnam-styles-invisible-horse-dance-craze#giddyup/2012/08/23/0dec6ace-ed29-11e1-b09d-07d971dee30a_blog.html.

54 Raj. “Why Gangnam Style Became So Popular ?? what is reason ?”, *Yahoo Answers*, <http://answers.yahoo.com/question/index?qid=20130117232246AAllg2t>.

lonfore. “What Makes GangNam Style so Popular???? I Cant See Anything Which is ExtraOrdinary!!!”, *Warrior Forum*, 29 May 2013, <http://www.warriorforum.com/off-topic-forum/795174-what-makes-gangnam-style-so-popular-i-cant-see-anything-extraordinary.html>.

Anonymous Coward. “How is Gangnam style so popular?”, *Godlike Productions*, 21 December 2012, <http://www.godlikeproductions.com/forum1/message2090536/pg1>.

Pegasus910601. “Why is gingham style so popular? (I don’t like it period)”, *Ask*, <http://www.ask.com/answers/288638101/why-is-gangnam-style-so-popular-i-don-t-like-it-period>.

scoutthesoldier. “Why is Gangnam Style so popular”, *Newgrounds*, 8 December 2012, <http://www.newgrounds.com/bbs/topic/1326318/3>.

7thbrah. “Why is Gangnam Style So Popular? I don’t get it.”, *Bodybuilding.com*, 29 September 2012, <http://forum.bodybuilding.com/showthread.php?t=148544483>

refers to a lavish lifestyle, characterised by people who belong to the upper class and are recognisable by a distinct appearance.⁵⁵

Gangnam Style belongs to the K-Pop (Korean popular music) genre. To understand some of the video's aesthetics, it is important to consider the nature of the K-Pop genre in comparison to American and European music video standards. K-Pop combines different popular music genres, e.g., pop, rock, hip-hop and R&B, and emphasises visual presentation, while blending Western and Asian values⁵⁶ (Shim 2006). K-Pop videos rely on carefully orchestrated and abundant imagery. A typical feature, also prominent in the *Gangnam Style* video, is the appearance of celebrity cameos. For those well acquainted with this stylistic framework, the video may not be particularly unusual. In fact, Psy has been well known in Korea before the release of *Gangnam Style*. But it is the unfamiliarity with its stylistic context that can be seen as part of *Gangnam Style*'s international success since audiences outside Korea, especially in Western countries, are neither familiar with the aesthetic context nor do they understand the lyrics. All that is left is a seemingly random sequence of images, a funny dance and the song's characteristic hook.

Shifman (2014) claimed that the *Gangnam Style* music video displays every sign of a typical meme. First, Psy appears like an ordinary person to audiences who are not familiar with Korean popular culture. He is displayed in a humorous light, which is characterised by his "flawed masculinity" (Shifman 2014, 85). The repetition of several elements is also important, such as the strong highlighting of the hook and the editing patterns that emphasise the dance scenes.

5.2.1.1 Content and Editing

The video shows Psy in various typical locations of the Gangnam district in Seoul. It does not follow a clear linear storyline; we first see Psy on a playground, followed by stables, then on top of a tall building, in a garage, a sauna, etc.

55 Steinmetz, Katy. "Wednesday Words: Gangnam Style, Geek Speak and More", *TIME*, 12 September 2012, <http://newsfeed.time.com/2012/09/12/wednesday-words-gangnam-style-geek-speak-and-more/>.

56 Benjamin, Jeff. "The 10 K-Pop Groups Most Likely to Break in America", *Rolling Stone*, 18 May 2012, <http://www.rollingstone.com/music/lists/the-10-k-pop-groups-most-likely-to-break-in-america-20120518>.

The appearance of five celebrity cameos should be noted that play a key role in the overall context and for the understanding of the video. They include the dancing boy, Hwang Min-woo, the man dressed in a yellow suit, emcee Yoo Jae-suk, the man dancing in the elevator, Noh Hongchul, the old men playing a board game, Daesung and Seungri, and finally Hyuna, the dancing girl who attracts Psy's attention in the subway (Figure 5.2.3). It is easy to mistake these appearances for random scenes that were placed in the video without particular meaning. However, considering that cameos are a typical K-Pop characteristic, we can easily observe that these celebrities were placed strategically throughout the video to spark further interest and curiosity throughout the video.



Figure 5.2.3: Celebrity cameos in "Gangnam Style" (officialPsy, "PSY - GANGNAM STYLE (쩔어) M/V, <https://www.youtube.com/watch?v=9bZkp7q19f0>).

The video does not exhibit a clear pattern of events which can be considered a story with a beginning, middle and end; neither is there an obvious reason for or a meaning in the succession of the various locations and profilmic events. The video does, however, display editing patterns that correspond strongly to the song's sonic elements, thereby underlining the beat, the metrical rhythm, the song structure and the lyrical content. However, these patterns are not applied consistently throughout the entire video.

As already mentioned, the most common motif is the “invisible horse dance”, which is always shown in long or medium shots (Figure 5.2.4). Close-ups are used sparingly in this video, i.e., only at a few points throughout the video to mark specific events, for instance, as a starting point to gradually reveal the rest of the scene or to highlight characteristic features of the song. The use of the close-up in this video is in line with Vernallis' (2004, 110) description of its

function as an element used to highlight “an intensification in the music” and as an element to stimulate memory in a “flow of images that are too rich and materials that seem to dissolve too quickly” (Vernallis 2004, 48).



Figure 5.2.4: The invisible horse dance (officialPsy, "PSY - GANGNAM STYLE (□ □ □ □ □) M/V, <https://www.youtube.com/watch?v=9bZkp7q19f0>).

Most noticeable is the highlighting of the hook “Oppan Gangnam Style”. The phrase is usually preceded by a short sequence of two rapidly juxtaposed shots in the pre-chorus segment that build tension. When exclaiming the above phrase, Psy is shown either in a close-up or in a wider shot that is supported by other salient visual features, such as an explosion in the background. The chorus then either starts with the revelation of a humorous setting or the performance of the invisible horse dance. Vernallis (2004, 156) also mentioned this pattern as being typical of music video, since a “song’s hooks are most likely to receive emphasis”, especially if they describe simple physical actions. “Oppan Gangnam Style”, in this case, prompts the protagonists and therefore the potential audience to dance.

The transition of musical phrases is reflected by the clear visual divisions in the video. As mentioned above, this confirms Vernallis’ (2004, 163) suggestion that music videos “will often draw attention to a song’s sectional divisions in order to help the viewer remember the song and its structure”. She further explained that this is usually achieved, for example, by using striking visual changes, by switching location or dominant colouring. In *Gangnam Style*, the beginnings and endings of musical phrases are either marked by the conclusion of a shown location or the introduction of a new location. The bridge, as the final build-up to the hook “Oppan Gangnam Style”, and subsequently the chorus display a similar pattern as seen in the foregoing pre-choruses, but tend to syncopate as song and video eventually prepare for conclusion. Cuts occur on the beat and on characteristic sounds more often in the beginning, and become less sound-driven towards the end of the song. The video content itself, however, remains highly rhythmic and synchronised largely by displaying dancing scenes.

5.2.1.2 Structure

- Intro A close-up of PSY shows him lying in the sun. The viewer is made to believe that he is at the beach, before the camera moves into a wider shot to reveal the location as a playground.
- Verse 1 PSY is shown dancing, singing and walking towards the camera with attractive women by his side. Hwang Min-woo appears as the dancing boy. Shot locations in this segment are: stables, a playground, the top of a large building, a warehouse/garage.
- Pre chorus 1 PSY is shown dancing in several humorous scenes in the following locations: a sauna, a park, a sports hall and a bus. Daesung and Seungri appear as old men playing a board game in a park. The segment prepares for conclusion with fast editing between two scenes and ends with PSY walking towards the camera and away from an explosion in the background.
- Chorus 1 All the scenes in the first chorus show the invisible horse dance. The end of this segment shows PSY yelling excitedly upon watching a woman exercising. Locations in this segment are a sports hall, an inner city setting, a riding hall, a carousel and a boat on a river.
- Verse 2 The second verse shows PSY in a garage in what appears to be a confrontation or “dance battle” with celebrity emcee Yoo Jae-suk. This is followed by a scene showing PSY with Noh Hongchul performing a peculiar dance in an elevator.
- Pre chorus 2 In this segment, PSY is shown approaching celebrity Hyuna in a subway train and station.
- Chorus 2 The scenes in the second chorus show the invisible horse dance and PSY diving and/or drowning in a pool. Locations in this segment are: a subway station, a pool, a street and the previously shown riding hall.
- Bridge The bridge starts with a distinctly different scene, a dark dance hall that is illuminated by bright colourful spot and laser lights. At the end of this section, PSY is shown in a close-up. As he exclaims the hook, the camera moves out again and reveals his location to be a bathroom.
- Chorus 3 The final chorus first cuts to the previously shown bus scene, then back to the dance hall.

Outro	Two previously used scenes are shown again: the dance battle in the garage with emcee Yoo Jae-suk and PSY walking towards the camera and away from an explosion in the park.
-------	--

5.2.2 Call Me Maybe

Carly Rae Jepsen's single *Call Me Maybe*⁵⁷ was released in September 2011. After several months the song quickly became successful after pop singer Justin Bieber endorsed it by first tweeting about it in December 2011⁵⁸ and then published a lip dub video with fellow celebrities Ashley Tisdale and Selena Gomez on YouTube in February 2012 (cp. Figure 5.2.5).⁵⁹



Figure 5.2.5: Justin Bieber, Twitter post (Justin Bieber, Twitter post, <https://twitter.com/justinbieber/status/152843702790914050>).

The celebrities' online endorsement of the, until then, largely unknown singer Carly Rae Jepsen was a valuable push for the musician allowing her to significantly increase the target audience's awareness. The song's popularity quickly gained momentum, and eventually, by the summer of 2012, the song topped various music charts worldwide. At the time of this writing, the music video for *Call Me Maybe* counts over 600 million views on YouTube.

Although the music video itself played a lesser role in promoting this song when compared, for instance, to *Gangnam Style*, YouTube, among other online platforms, did still play a vital

57 CarlyRaeJepsenVEVO. "Carly Rae Jepsen - Call Me Maybe", *YouTube* video, 3:19, 1 March 2012, <https://www.youtube.com/watch?v=fWNaR-rxAic>.

58 Billy, Matthew. "How Powerful is Justin Bieber's Twitter Account?", *hypebot.com*, December 2012, <http://www.hypebot.com/hypebot/2012/12/how-powerful-is-justin-biebers-twitter-account.html>.

Lau, Melody. "Justin Bieber Gives Singer Carly Rae Jepsen", *Rolling Stone*, 12 March 2012, <http://www.rollingstone.com/music/news/justin-bieber-gives-singer-carly-rae-jepsen-a-boost-20120312>.

59 CarlosPenaTV. ""Call Me Maybe" by Carly Rae Jepsen - Feat. Justin Bieber, Selena, Ashley Tisdale & MORE!", *YouTube* video, 3:25, 18 February 2012, <https://www.youtube.com/watch?v=AsBsBU3vn6M>.

role in the song's promotion. In addition to the massive number of views of the original clip, throughout 2012 a variety of parodies, remixes and lip dubs were uploaded by fans and celebrities alike, turning "Call Me Maybe" into a full-fledged meme that permeated social media for months.⁶⁰

A large number of people have attributed the song's popularity to its catchy hook and memorable chorus.⁶¹ The lyrics to this up-beat song are simple and repetitive, and therefore easy to learn. Eventually Billboard announced *Call Me Maybe* the "Song of the Summer in 2012".⁶² It was also nominated for two Grammy awards⁶³ and three MuchMusic Video Awards,⁶⁴ among others.

5.2.2.1 Content and Editing

The video follows a clear and simple humorous narrative combined with performance elements that are embedded in the storyline. The story progresses in the following stages:

- 1) Jepsen watches the video's protagonist, who is presented as her neighbour through the window of her house,
- 2) Jepsen, together with her band, performs the song in her garage,

60 minigreek87. "Call Me Maybe", *Know Your Meme*, <http://knowyourmeme.com/memes/call-me-maybe>.

61 Sher, Lauren. "Why 'Call Me Maybe' Is So Catchy", *ABC News*, 22 Jun 2012, <http://abcnews.go.com/blogs/entertainment/2012/06/why-call-me-maybe-is-so-catchy/>.

Caporimo, Alison. "'Call Me Maybe': Why Is It So Addictive?", *Readers Digest*, <http://www.rd.com/recommends/call-me-maybe-why-is-it-so-addictive/>.

Freeman, Michael. "Explainer: Why is Call Me Maybe so very, very catchy?", *The Daily Edge*, 21 July 2012, <http://www.dailymotion.ie/why-call-me-maybe-so-catchy-explainer-527715-Jul2012/>.

Forbes, Lisa. "'Call Me Maybe' Offers Blueprint For Marketing Success", *Forbes*, 7 December 2012, <http://www.forbes.com/sites/lisaarthur/2012/07/12/call-me-maybe-offers-blueprint-for-marketing-success/>.

62 Trust, Gary. "Carly Rae Jepsen's 'Call Me Maybe' Is Billboard's Song of the Summer", *Billboard*, 6 September 2012, <http://www.billboard.com/articles/news/475358/carly-rae-jepsens-call-me-maybe-is-billboards-song-of-the-summer>.

63 Hernandez, Brian Anthony. "'Call Me Maybe' is Nominated for Grammy's Song of the Year", *Mashable*, 6 December 2012, <http://mashable.com/2012/12/06/grammy-nominations-call-me-maybe-carly-rae-jepsen/>.

64 Flare Entertainment Team. "2012 Muchmusic Video Award Nominees Released", *Flare*, 16 May 2012, <http://www.flare.com/celebrity/2012-muchmusic-video-award-nominees-released/>.

- 3) Jepsen's band mates encourage her to attract her neighbour's attention while he is repairing his car next door. Following this she conspicuously washes her car,
- 4) she falls off her car and dreams about being the couple on the cover of a romantic novel,
- 5) the neighbour helps her,
- 6) Jepsen and the band play the song for him,
- 7) the neighbour gives his number to one of her band mates.

There are no distinctly noticeable editing patterns in this video. However, a sense of rhythm is created by cuts frequently following the song's beat, especially during the first half of the song. The song's musical phrases are highlighted to some extent, but not consistently throughout the video. Verses generally tend to focus more on the story, while the choruses dominantly show performance. In this video, performance is used as both a progressive narrative element as well as a stylistic extension that showcases the singer and other musicians (Figure 5.2.6).



Figure 5.2.6: Screenshots of the *Call Me Maybe* music video (CarlyRaeJepsenVEVO, Carly Rae Jepsen - *Call Me Maybe*, <https://www.youtube.com/watch?v=fWNaR-rxAic>).

A analysis of the pre-chorus and the chorus reveals a subtle editing pattern. The pre-choruses have an average shot duration of about 1.5 seconds, which is shorter than the average shot durations in the rest of the video. This results in a faster editing pace and rhythm, which supports the creation of tension just before the beginning of the chorus, thereby highlighting the most memorable sections of the song. While this pattern is less pronounced in *Call Me Maybe*, it is similar to the technique used in *Gangnam Style* to highlight the chorus.

The visual aesthetics regarding setting and colouring as well as rhythm and the overall pace remain comparatively consistent, except for the bridge section. The story takes place in and around Jepsen's house, focusing solely on the events around her home and her experience with her neighbour. The visual appearance of the images could be described as "vintage", with a colour treatment reminiscent of, for instance, Instagram photos.

A noticeable structure can be observed when watching the video with the music's sectional segmentation in mind, which helps to break up the story into individual scenes. One dramatic change in the overall visual appearance occurs during the song's bridge. Here, the video switches to a dream scene resembling the cover of a romantic novel. In this video segment, the location, colour scheme and the appearance of both Jepsen and the video's male protagonist are subject to a notable visual transformation.

The video's visual content features a wide range of different camera distances and movement. These technical parameters barely seem to follow the song's structure and are used primarily from a narrative point of view. Compared to *Gangnam Style*, there is little use of the wide shot, apart from establishing and maintaining scene and narrative orientation. Close-ups are frequently used, especially to highlight Jepsen as the vocalist during performance scenes.

5.2.2.2 Structure

- | | |
|--------------|--|
| Verse 1 | Jepsen is seen watching her neighbour mowing the lawn through her window. |
| Pre-Chorus 1 | The video sequence builds up tension with a slight increase in editing pace as she is watching him. |
| Chorus 1 | On the transition from pre-chorus to chorus, the male protagonist appears to notice Jepsen and as the chorus begins she drops down to hide under the window. Following this event, the band is shown playing the song in her garage. The scene cuts between this performance and the male protagonist mowing the lawn. |
| Verse 2 | Jepsen's neighbour is in front of his house, repairing his car. Her band mates convince her to attract his attention by washing her own car at the same time. |
| Pre-Chorus 2 | The beginning of the second pre-chorus cuts back to the band playing in the garage and then shows Jepsen's failed attempts in getting her neighbour's attention. |
| Chorus 2 | The section continues to show performance elements as well as Jepsen washing her car. |
| Bridge | As the male protagonist notices Jepsen, she falls off her car, losing consciousness. She dreams of the two of them being the couple on the cover of a romantic novel. |
| Chorus 3 | In the final chorus, the video cuts back to Jepsen lying on the ground waking up after her neighbour came to her aid. The band performs the song for |

Jepsen's neighbour until they finish and Jepsen goes to write down her number. The song ends with Jepsen's neighbour giving his number to one of her band mates instead.

5.2.3 Big Jet Plane

The song *Big Jet Plane*⁶⁵ was released as a single by the Australian duo Angus & Julia Stone. It was awarded an ARIA Music Award⁶⁶ and voted number 1 by listeners on the annual *Hottest 100* by Australian radio station Triple J,⁶⁷ a listener poll conducted by the Australian ABC. The *Hottest 100 2010* are also used as the basis for the last music video analysis in this section. The song was featured in several American TV series and also in the French film *Les Emotifs Anonymes (Romantics Anonymous, 2010)*. At the time of this writing, *Big Jet Plane* has over twenty million views on YouTube.

The video shows a young woman going about her daily routines as a shop assistant while she dreams of travelling. The song can be classified by popular music genres such as indie, folk and pop. Generally, Angus & Julia Stone's music is characterised by low-tempo simplistic acoustic compositions. Their music videos often tend to reflect this musical style with the use of a low-pace editing and a focus on emotional expression.

While *Big Jet Plane* was neither a worldwide viral success nor a pop cultural phenomenon like *Gangnam Style* and *Call Me Maybe*, the video was included in this series of case studies specifically as an example of a non-viral, but successful song and music video that belongs to a different genre of popular culture and music than the other case studies. The marketing for this song followed traditional paths and did not make particular use of social media.

Compared to Psy and Carly Rae Jepsen, Angus & Julia Stone have been well known in Australia, but to a lesser degree internationally in the years before the release of *Big Jet Plane*. In fact, the song was part of their second album, which was released in 2010. *Big Jet Plane* was therefore not a breakthrough or otherwise unusual phenomenon in popular music, but it was the duos biggest success up to that point.

65 NettwerkMusic. "Angus and Julia Stone - Big Jet Plane [Official Music Video]", *YouTube* video, 4:05, 30 March 2014, <https://www.youtube.com/watch?v=yFTvbcNhEgc>.

66 ARIA Awards. "History", *ARIA Awards*, <http://www.ariaawards.com.au/history/year/2013>.

67 ABC. "Hottest 100", *ABC*, <http://www.abc.net.au/triplej/hottest100/10/>.

5.2.3.1 Content and Editing

The video tells a simple story, with a young girl as the protagonist. She is first seen in her working environment as a shop assistant. These visuals generally evoke a feeling of dullness and boredom, as the viewer follows her monotonous actions. Eventually, the protagonist is shown as she finishes work for a day, followed by a scene of her spending the evening in her living room, where the viewer is made to associate the protagonist's boredom with a wish to travel. After this scene, the video shows the protagonist being back in the shop. This time the scene intercuts between the shop location and the protagonist being in a car, surrounded by nature.

The video starts with a close-up shot of the protagonist's hair in the wind and then cuts straight to her standing in front of a fan at the cashier's desk of her workplace (Figure 5.2.7). This first shot juxtaposition (which is purely based on one connecting element – the wind in her hair) sets the stage for the rest of the video, which revolves around symbolic images and the introverted nature of this simple narrative.



Figure 5.2.7: Shot juxtaposition using wind as connecting element (NettwerkMusic, Angus and Julia Stone – Big Jet Plane [Official Music Video], <https://www.youtube.com/watch?v=yFTvbcNhEgc>).

The video can be divided into three visually distinct sections, which are broken up by the song's bridge (Figure 5.2.8). In the first section, the protagonist is seen performing various activities in the shop, e.g., restocking products on the shelves, wandering the aisles and, finally, cleaning up after a long day of work. The last shot of this section shows the protagonist in front of her locker where she takes a last glance at a postcard stuck on the inside of the locker door.



Figure 5.2.8: Screenshots of the *Big Jet Plane* music video (NettwerkMusic, Angus and Julia Stone - *Big Jet Plane* [Official Music Video], <https://www.youtube.com/watch?v=yFTvbcNhEgc>).

After the start of the bridge, she is seen at home watching TV. This section introduces the only footage of the band, which can be seen playing the song on TV. The protagonist's living room is filled with symbolic imagery of travelling: a world map, vintage looking travel brochures, and postcards. Whereas in the previous section the viewer was made to observe her dull day at work, this section intensifies in emotion, ranging from dullness and perhaps sadness to longing and desire, by strongly implying her wish to travel with the juxtaposition of these symbols, predominantly displayed in close-ups.

At the transition between bridge and final chorus, the video first shows an extreme close up of the protagonist's eyes in a stroboscopic light, thus producing a similar effect to the rapid intercutting in the *Gangnam Style* video, which is used to build up tension before eventually returning to the chorus. As the chorus starts, the protagonist is shown in the shop and the sequence intercuts between this location and shots of her in a car surrounded by nature, which is presumably the same location shown in the very first shot.

Of the three videos analysed in this section, *Big Jet Plane* most strongly reflects Vernallis' (2004, 195) description of loose synchronisation and overall connections between shots becoming "more tenuous", "simple", and "broadly spaced" as opposed to openings and middles where the editing will tend to reflect the music more closely.

Compared to the previous two videos, due to the song's and video's slow tempo, the visual rhythm plays a minor role. Although the video also contains less narrative events than *Call Me Maybe* in particular, the editing is more strongly oriented towards the simple narrative shown in this video, largely by focusing on the protagonist's expression and details of the locations' mise en scène, which establish the protagonist's context and serve to match the musical tone. The editing follows the song's form at some moments of musical tension but does not apply consistent audio-visual synchronisation. Particularly, while the bridge serves to break up the video into three visually distinct sections, the editing does not strictly follow the musical segmentation.

5.2.3.2 Structure

- | | |
|----------|--|
| Intro | A shot of the protagonist's hair in the wind, followed by a shot focused on her standing in front of a fan in the shop. She is seen scanning an item, waiting, and restocking products. |
| Chorus 1 | Two wide shots establish the shop location with the protagonist stocking shelves and wandering down the aisles. |
| Verse 1 | Three shots show the protagonist carrying boxes. |
| Chorus 2 | The protagonist throws out the spare boxes and goes to her locker where she glances at the postcard on the locker door and eventually closes the door. In the next shot she is seen in her living room, with a strong colour contrast from the natural lighting in the shop to a deep blue tint. She watches the band perform the song on her TV while several shots show her surroundings, e.g., posters, postcards and brochures symbolising travelling. |
| Bridge | Close-ups of the TV show the band playing the song. A close-up of the protagonist shows her looking at her hands, followed by an extreme close-up of her eyes. |
| Chorus 3 | A sequence of shots intercutting between the protagonist in the shop and in a car is shown until eventually the protagonist is shown standing in front of the shop's sliding door, cleaning the glass. |

5.3 Interactive Music Videos

In an early examination of automated or “virtual” video editing, Mackay and Davenport (1989, 804) described film and video production as a process that is “highly interactive during the making”, but that “quite intentionally minimises interaction on the part of the audience in order to allow viewers to enter a state of reverie”. Dancyger (2011, 70) suggested that film seeks to create an experience in which viewers will be “very involved and not involved at all”. Watching movies and television is generally considered a way of “passive” and “relaxing” media consumption (Vorderer 2000, 71, Rieser & Zapp 2002, Viola 2003). With regard to television, studies suggest that the medium is used for a number of purposes, for example to escape everyday life, to receive information on current events and to compare one’s life to those of others (Condry 1989). Lee and Lee (1995) also observed that levels of audience

attention can vary greatly. Many viewers keep the television turned on while engaging in other activities “like an electric light” (Lee and Lee 1995, 11).

While interactive TV and cinema have been anticipated and experimented with for many years, especially with regard to viewer influence on narrative progression (Szilas 1999, Kim 2001, Boddy 2004, Ben Shaul 2008, Chorianopoulos and Lekakos 2008, Barker 2012), it seems unlikely that these types of media formats will become widely adopted within their established media environments, i.e., television and movie theatre, because they abandon the aforementioned concept of passive engagement and reverie, by trying to turn the viewer into an active participant who can influence a given program, either to explore alternative narrative routes, or simply to adapt a story to their personal preferences.

An area which has always sought to combine moving images with interactivity is video art (Wardrip-Fruin and Montfort 2003). Additionally, online video, for instance distributed on platforms such as YouTube, has been established within the context of interactive interfaces, which maximise viewer control over playback.

Traditional narratives, as well as the relationship between media, author and audience, seem to be changing with increasing incorporation of digital media, leading to what Manovich has called the database form, a concept which he would later develop into his definition of software-based media (Manovich 2001, Rieser and Zapp 2002, Manovich 2013). Dancyger (2011, 394) suggested that “the digital age demands new narrative styles for the new but fragmented audience of the age”. According to Dancyger (2011, 395), non-linear narratives are “intuitive rather than purposeful, random rather than developmental” and emphasise “feeling over action”. Video formats are changing accordingly and new technologies are used to test online video’s capabilities and possibilities.

In order to illustrate some of the innovation surrounding online video, three examples of interactive music videos will be discussed in the following subsections, i.e., *The Wilderness Downtown*,⁶⁸ *Like a Rolling Stone*,⁶⁹ and *Do Not Touch*.⁷⁰ As mentioned in Chapter 2, many interactive or non-linear music videos are based on audience participation, either before or

68 The Wilderness Downtown. <http://www.thewildernessdowntown.com/>.

69 Bob Dylan. “Bob Dylan “Like a Rolling Stone””, <http://video.bobdylan.com/desktop.html>.

70 Do Not Touch. <http://donottouch.org/>.

during video playback. Hence, the overall progression as well as the user experience created by these videos can be based on different kinds of interactivity, from preliminary input to actions during playback with keystrokes, mouse clicks and other movements on the graphical interface.

Hammoud (2006, 5) defined several different types of interactive video. The type of interactive music videos discussed here falls into Hammoud's category of the "interactive video presentation" type, which he described as "a form of interactive video document that is centred on enriched video but is not exclusively video". Alternatively, Hammoud (2006, 6) denoted these videos "hyperfilm", "hypervideo", or, more generally, "hypermedia".

In recent years, interactive music videos seem to have established a field of their own. While some authors whether the terms "music video" or even just "video" are applicable to these moving-image experiences on the Internet (Edmond 2012), increasingly the term "interactive music video" is used to refer to any music video that isn't strictly linear and includes some sort of interactivity.⁷¹

Initially, interactive music videos were created based on Adobe Flash (previously Macromedia Flash and Shockwave Flash), which had been the standard solution for any online application of animated graphics, interactive multimedia applications and, since 2003, online video streaming. An early example for an interactive music video using Flash is *Neon Bible* by Arcade Fire.⁷² While Flash is still widely used for many implementations of online video streaming and interactive applications, for example, Interlude's Treehouse software, the use of Flash has declined since the new HTML5 standard allowed for the native implementation of video content and similar modes of implementing animated graphics and interactive multimedia applications (and since Apple's late Steve Jobs publicly criticised Flash for being unsuitable for mobile devices,⁷³ with the result that compatibility with Flash was removed from Apple's mobile devices).

71 Lewis, Jennifer. "A Selection of Awesome Interactive Music Videos", *Flavorwire*, 24 February 2012. <http://flavorwire.com/261427/a-selection-of-awesome-interactive-music-videos>.

Gibsone, Harriet. "Top Five: Interactive Music Videos", *The Guardian*, 23 November 2013. <http://www.theguardian.com/music/musicblog/2013/nov/22/top-five-interactive-videos-bob-dylan>.

72 Beonlineb. <http://www.beonlineb.com>.

73 Jobs, Steve. "Thoughts on Flash", *Apple*, April 2010, <http://www.apple.com/hotnews/thoughts-on-flash/>.

In 2010, a number of interactive music videos were created to specifically use and show the capabilities of Google's new web browser Chrome in conjunction with HTML5. At that time, Google Chrome was gradually entering the web browser market and Chrome Experiments were launched in order to demonstrate its compatibility with various technologies. At the same time, HTML5 had slowly been published across all web browsers with varying compatibility. It was also around this time that Google purchased On2 along with the VP8 video format. Subsequently, Google launched WebM, which was meant to serve as a royalty-free video format for HTML5.⁷⁴ Therefore, for some time, Google linked Chrome to the viewing of interactive HTML5 applications, in this case, music videos, with the release of *The Wilderness Downtown*, *3 Dreams of Black*,⁷⁵ and *All Is Not Lost*,⁷⁶ among others.⁷⁷ Since the HTML5 standard has still been in active development and due to subsequent lacks in cross-browser compatibility as well as Google's marketing strategies, these videos could only be viewed properly in Chrome and therefore required many users to download the new browser if they wanted to watch them.

According to w3schools.com,⁷⁸ as of October 2014 Chrome has become the most widely used browser, with a current market share of 60.4%, and the HTML5 standard has officially been completed by the W3C.⁷⁹ While the use of HTML5 video and canvas allows for many possibilities in the creation of multimedia applications, Flash still offers several advantages over the new standard, for instance, with regard to cross-browser compatibility and stability. It therefore remains to be used on many popular platforms and applications.⁸⁰

74 WebM. "About WebM", <http://www.webmproject.org/about/>.

75 ROME. "3 Dreams of Black", <http://www.ro.me>.

76 All Is Not Lost. "All Is Not Lost", <http://www.allisnotlo.st>.

77 Chrome Experiments. "Experiments: Interactive Music Video", <http://www.chromeexperiments.com/tag/interactive-music-video/>.

78 w3schools. "Browser Statistics", http://www.w3schools.com/browsers/browsers_stats.asp.

79 Cotton, Paul. "W3C Charts a Course for the Future of the Open Web by Finalizing HTML5 Standard", *Microsoft Open Technologies*, 28 October 2014, <http://msopentech.com/blog/2014/10/28/w3c-charts-a-course-for-the-future-of-the-open-web-by-finalizing-html5-standard/>.

80 W3Techs. "Usage of Flash for Websites", <http://w3techs.com/technologies/details/cp-flash/all/all>. Alex. "Why we still use Flash™", Medium, 11 October 2014, https://medium.com/@h_a_l_e_x/why-we-still-use-flash-104b8bf236ca.

Zynga Player Support. "Guide to Installing Adobe Flash Player", *Zynga Player Support*, 2 December 2014, https://www.zyngaplayersupport.com/article/farmville-2/Guide-to-Installing-Adobe-Flash-Player-en_US. Mittal, Romil. "AIR app installs cross a billion", *Adobe AIR and Adobe Flash Player Team Blog*, 20 April 2014, <http://blogs.adobe.com/flashplayer/2014/04/air-app-installs-cross-a-billion.html>.

While two of the interactive videos discussed in this section, i.e., *The Wilderness Downtown* and *Do Not Touch*, were created using HTML5, Interlude's software *Treehouse*, as mentioned above, is still based on Flash. This, however, has not stopped the company from drawing attention towards its interactive music videos again in 2013 and 2014. *Like a Rolling Stone* had, at the time of this writing, become Interlude's newest prime example for the software's creative capabilities.

While videos created with *Treehouse* come closest to creating a new type of standard for the use of interactivity and non-linear methods for video creation, the other two videos discussed in this section provide completely different approaches to user engagement and technological experimentation. In addition to different modes of interactivity, these videos also cover a range of different aesthetic concepts. The following case studies will highlight how these videos seek to engage the viewer and the methods used in this process.

Although linear music videos, e.g., the videos examined in the previous section, will likely remain essential in online media, successful non-linear music videos are treated with a completely new interest that focuses as much on the video and its author as on the song and the musician that it promotes. Each of the following videos was mentioned in countless news outlets and technology magazine articles, for instance, interviews with the directors and the creative teams discussing video concepts, ideas and motivations. In these cases, as a matter of fact, by studying the media coverage and promotion campaigns, we will learn more about the video than the featured songs.

However, in this context, it is also necessary to consider the role of innovation itself, instead of focusing merely on the impact of creative outcomes. Linear music video is an established media form. Even those videos that can be considered highly successful may not be highly innovative. Their effectiveness may largely be rooted in their use of established aesthetics and conventions. Therefore, it is necessary to examine how the following interactive music videos generate user engagement, and to be mindful of the extent to which these methods are simply showcasing technical innovation rather than creating methods and frameworks that can be applied in a larger number of videos.

5.3.1 The Wilderness Downtown

*The Wilderness Downtown*⁸¹ is an interactive music video created by Google and Chris Milk for the band Arcade Fire and their song *We Used to Wait*. According to Milk, the idea for creating *The Wilderness Downtown* can be traced back to Google's suggestion to use a media project to demonstrate HTML5's capabilities.⁸²

In this video, the protagonist, a neutral figure dressed in a hooded sweater, is seen running on a street (Figure 5.3.1). The video's visual content depends on the prior input of an address, which triggers the site to fetch images from Google Maps and Street View. The overall look and feel of the video footage is kept simple in order to support the viewers' personal association and imagination.

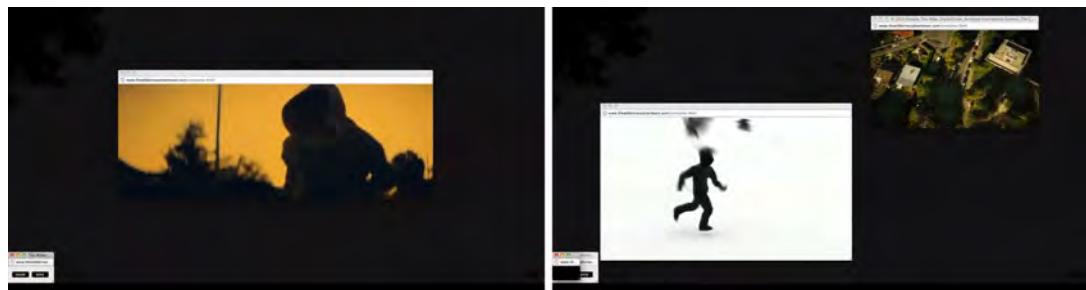


Figure 5.3.1: Screenshots of the *The Wilderness Downtown* music video (*The Wilderness Downtown*, <http://www.thewildernessdowntown.com>).

Although the video's narrative and visual progression are comparatively simplistic, its popularity can largely be ascribed to the highly innovative concept for online video presentation and the use of upcoming technologies.⁸³ Characteristic in this sense is the use of

81 The Wilderness Downtown. <http://www.thewildernessdowntown.com/>.

82 Castillo, Michelle. "The Wilderness Downtown's Creator Talks About What Motivated Him, What's Next", *TIME*, 20 September 2010, <http://techland.time.com/2010/09/10/the-wilderness-downtowns-creator-talks-about-what-motivated-him-whats-next/>.

83 eBiz MBA. "Top 10 Best HTML5 Sites of 2015", *eBiz MBA*, <http://www.ebizmba.com/articles/best-html5-websites>.

Ehrlich, Brenna. "Arcade Fire's Experimental New Video Shows What's Possible with HTML5", *Mashable*, 31 August 2010, <http://mashable.com/2010/08/30/arcade-fires-chrome-video/>.

Heywood, Matt. "Google Shows Off Some HTML5 Magic With "The Wilderness Downtown" Experiment", *Entertainment Buddha*, 30 August 2010, <http://mashable.com/2010/08/30/arcade-fires-chrome-video/>.

Bent, Meaghan. "Arcade Fire + Google Maps = The Wilderness Downtown", *IndieBlog*, 31 August 2010, <http://mashable.com/2010/08/30/arcade-fires-chrome-video/>.

pop-up windows to introduce new shots and scenes, rather than depicting the entire video within its one rectangular field where it is usually made to reside.

Accordingly, Milk described the editing process as follows:

*"The editorial process was really interesting. A cut/edit has a certain weight to it. You use that weight as a tool. But a whole new browser window popping up with a new film element in it - that has its own percussion and tone. We had to edit differently than we normally would. It's a different canvas than a television or a cinema screen. It's more comparable to plasma screens popping up all over your wall - needless to say, it's an effect you need to be a little careful with."*⁸⁴

The technique is reminiscent of video installations pioneered by Nam June Paik (1993).

A central concept in this music video is to provide a personal viewing experience while showcasing technology. On the one hand, this is achieved by adding potentially familiar images, which are combined with other video elements, such as the running protagonist, a flock of birds, and growing trees. On the other hand, the user has the option to interact with the video in real-time, by clicking, writing, and drawing at particular points. Milk explained that, while he enjoyed experimenting with technology, his primary concern was to create something "that would emotionally resonate with people" and to maintain a work's "humanity". Google maps and Street View are, he further elaborated, "cold high-tech that can be incredibly emotional when used in the right context".⁸⁵

The Wilderness Downtown introduced new concepts for viewing video and has even been discussed in the context of architecture and theatre. Lavin (2011, 400) described the experience of viewing this "film" from the moment she encountered the interface where she was asked to input information and hit play. Irwin (2011, 57) described *The Wilderness Downtown* as an "ephemeral performance" that creates a new sense of audience engagement and participation.

Houk, Justin. "Arcade Fire Video Uses Google Maps to Bring You Home", ProgrammableWeb, 6 September 2010, <http://www.programmableweb.com/news/arcade-fire-video-uses-google-maps-to-bring-you-home/2010/09/06>.

84 Castillo, Michelle. "The Wilderness Downtown's Creator Talks About What Motivated Him, What's Next", TIME, 20 September 2010, <http://techland.time.com/2010/09/10/the-wilderness-downtowns-creator-talks-about-what-motivated-him-whats-next/>.

85 ibid.

5.3.1.1 Content and Editing

First, the user is asked to provide an address, preferably his or her first home address, before being able to play the video. This address is then used to fetch images from Google Maps and Street View, which are shown during playback. The video is displayed and divided among an alternating number of individual windows, which pop up and disappear on the screen at certain points during the sequence. Possibilities for interaction during the video include mouse movements to alter the motion of a flock of birds, as well as mouse and text input for the purpose of writing a virtual postcard, which can be sent by email at the end of playback.

The entire video shows the protagonist running down a street. The editing of video sequences within individual windows is highly simplistic and relies on displaying shots of very long durations. A stronger sense of editing for the fabrication of meaning arises through showing, combining and moving windows that contain video footage, animation or images from Google Maps and Street View. In this sense, it could be said that the video provides its own mode of editing to function in this technological context.

The *Wilderness Downtown* hardly follows a narrative structure but primarily aims to create an emotional experience through the use of music and symbolic imagery. In this way, it can also serve as an interesting example for Dancyger's (2011, 168) definition of "MTV style", which, according to him, is based on the creation of "a definite feeling state" and the abandonment of traditional narrative in audio-visual works.

The overall editing and composition of the video can be considered as slow-paced throughout its entire duration. Visual rhythm is largely created via the protagonist's running motion in conjunction with the simple piano melody, with a sense of synchronisation emerging when the two are isolated. Lavin (2011, 400) observed "the popping up of windows" as happening at "random times", which illustrates the unfamiliarity and notion of disunity that many viewers may experience upon watching *The Wilderness Downtown*. Although the video's visual composition can generally be considered as unusual, we can observe that the movement, appearance and disappearance of windows usually follows the musical beat as well as characteristic moments in the song, often on sectional transitions.

While a similar approach was taken for a following Chrome Experiment, called *All is Not Lost*⁸⁶ based on a song by OK Go, the method of displaying varying video sequences simultaneously in a number of different windows has not been replicated in any major works. The greatest challenge for this method of video presentation might be practicality. Users are accustomed to multitasking when viewing online content, switching between browser tabs and other software applications. A music video like *The Wilderness Downtown*, however, requires a user's full attention, since any other activity on the screen would not only interfere with the intended experience of viewing the video but also compromise the actual display and functionality.

The technique of personalising video content was also used in OK Go's *All is Not Lost*, and in other videos such as the music video for Linkin Park's *Lost in the Echo*,⁸⁷ where the user is prompted to provide their Facebook details. It is likely that these types of personalisation techniques will be used more frequently in the future. As well as being utilised to elicit an emotional response from a viewer, whether nostalgic or humorous, these techniques can also serve to individualise other digital media forms, e.g., in advertising, entertainment and information, to increase attention, interest and engagement.

5.3.1.2 *Interactivity*

The interactivity and variability in *The Wilderness Downtown* are comparatively limited. The video's primary functionality lies in the preliminary provision of information by the viewer, in this case an address which is retrievable on Google maps and Street View. Additional methods of viewer interactivity include mouse movements, which influence the movement of a flock of birds and text input on a virtual postcard.

These interactive methods have in common that individual elements in the video can be influenced, but not the video's progression. This results in a high level of control from the author's point of view, while providing a visually customised video with some real-time interaction.

86 All Is Not Lost. "All Is Not Lost", <http://www.allisnotlo.st>.

87 Lost in the Echo. "Linkin Park | Lost in the Echo", <http://lostinthecho.com/>.

5.3.1.3 Structure

- Verse 1 One window pops up showing the running man, first his feet and then his upper body, with his face concealed by his sweater hood. After half a minute, a second window pops up showing a flock of birds in the sky whose motion can be influenced with the cursor.
- Verse 2 The previous images continue. At the end of the pre-chorus, the birds in the second window move down and at the same time this window changes both position and size, thereby turning into the new centre window and removing the first.
- Pre-Chorus 1 The centre window depicts a Google Street View, depending on the prior input, with the birds superimposed on the image. A second window appears on the left, showing the running man again in a portrait format. The window disappears and another one pops up on the bottom right, showing the running man in landscape mode. The window disappears again and a new one pops up at the top right, showing the running man in landscape format again.
- Verse 3 The centre window remains the same, but the top right window disappears again and a new window pops up in the bottom right corner, showing the running man's feet again. An additional window pops up on the left showing a Street View of the user's address. The window disappears for a while and then reappears on the right. Then, the window disappears and a new window pops up showing a full body shot of the running man on the left, in addition to the second window depicting only his legs. The last window disappears and is replaced with a window that shows the Street View again. At the end of this section, the right window moves to the centre, now showing that the running man has stopped, and the centre window moves to the left.
- Transition The protagonist and the Google Maps footage in the left window start to spin. A third window appears showing the Street View and spinning as well.
- Bridge The centre and right windows disappear and the left window moves to the centre, now showing a bird's eye view of the streets. At the end of the bridge, the window shrinks and disappears.
- Verse 4 A new window appears in the top left corner which lets the user write a text and draw on a virtual postcard. Three small windows depicting running machinery appear underneath.

- Pre-Chorus 2 No change.
- Chorus 2 No change until a new window pops up towards the end, showing a 3D-animated version of the running man on white space.
- Outro The last window disappears and birds fly in on the window depicting the postcard, landing on whatever the user has drawn onto it. With the return of the vocals, the window depicting the 3D runner pops up again on the right. A new window pops up with footage from Google Maps. The birds begin to fly again. As they seem to fly out of the postcard, they appear in both the Google Maps window and the running man window. All windows apart from the top right one showing the Google Maps footage disappear. A new window pops up on the bottom left showing the runner, flying birds and trees erupting as some of the birds dive into the ground. Windows showing these trees pop up and disappear on different locations on the screen, until the footage from Maps and Street View can be seen to be overgrown with trees. Finally, all windows disappear.

5.3.2 Like a Rolling Stone

In the interactive video for *Like a Rolling Stone*,⁸⁸ a graphical interface simulates a TV set with sixteen channels, with each showing a different program, e.g., the news, a game show, and a tennis match (Figure 5.3.2). Viewers have the option to watch any of these channels and switch between them throughout the entire duration of this interactive music video.



Figure 5.3.2: Screenshots showing three different channels of the *Like a Rolling Stone* interactive music video (Bob Dylan, *Like a Rolling Stone*, <http://video.bobdylan.com/>).

Directed by Vania Heymann, *Like a Rolling Stone* is, at the time of this writing, one of the latest products created with Interlude's interactive video software Treehouse and was marketed as the first ever music video for this 1965 song by Bob Dylan, as part of the marketing campaign for a new CD box set by the musician. According to Yoni Bloch, CEO of

88 Bob Dylan. "Bob Dylan 'Like a Rolling Stone'", <http://video.bobdylan.com/>.

Interlude, the video attracted almost one million unique viewers on the first day of its release, with over 90% of these viewers making use of the interactive features.⁸⁹

A large number of TV shows and articles on blogs, in magazines and various news outlets reported on the video's innovative presentation techniques for a song that many consider a cult classic.⁹⁰ In these overwhelmingly positive articles, the video was called, e.g., "the most impressive I've seen",⁹¹ "an interactive masterpiece",⁹² and "the most amazing music video you will see all year".⁹³

5.3.2.1 Content and editing

The concept for this interactive music video is very simple: sixteen television channels are provided which can be viewed at any point during the video for as long as the users wish to do so. This results in unique combinations of video content when the viewers decide to switch between these channels.

In this case, interactivity is therefore used to 1) create a different music video, every time it is accessed⁹⁴ and 2) to create a playful, exploratory, as well as humorous viewing experience. Due

89 Edwards, Gavin. "Inside Bob Dylan's Brilliant 'Like a Rolling Stone' Video", *Rolling Stone*, 20 November 2013, <http://www.rollingstone.com/music/news/inside-bob-dylans-brilliant-like-a-rolling-stone-video-20131120>.

90 Interlude. "How Does It Feel?", *Interlude* blog, 22 November 2013, <http://blog.interlude.fm/?p=106>.

91 Haglund, David. "This New Video for "Like a Rolling Stone" Is Fantastic", *Slate*, 19 November 2013, http://www.slate.com/blogs/browbeat/2013/11/19/like_a_rolling_stone_new_interactive_bob_dylan_video_by_vania_heymann_is.html.

92 Maloney, Devon. "The First-Ever Official Video for 'Like a Rolling Stone' Is An Interactive Masterpiece", *Wired*, 19 November 2013, <http://www.wired.com/2013/11/bob-dylan-rolling-stone-video/>.

93 Weaver, Caity. "Stop Reading This Sentence and Just Watch Bob Dylan's Incredible Video", *Gawker*, 19 November 2013, <http://gawker.com/stop-reading-this-sentence-and-just-watch-bob-dylans-i-1467612881>.

94 Pelly, Jenn. "Watch Bob Dylan's Insane Interactive "Like a Rolling Stone" Video, Starring Danny Brown", *Pitchfork*, 19 November 2013, <http://pitchfork.com/news/53063-watch-bob-dylans-insane-interactive-like-a-rolling-stone-video-starring-danny-brown/>.

to the large amount of video content available to viewers (one hour and fifteen minutes in total),⁹⁵ the entertaining value in repeated viewing can be effectively maintained.

Each video channel in this interactive music video constitutes a comparatively authentic imitation of real television shows. Every individual sequence is driven by Dylan's vocals. The protagonists' lip-synching to these vocals is generally foregrounded. While the vocals and visual representation sometimes seem to connect, most of the time they do not seem to share a connection. This can sometimes make the progression feel a little too slow and even dull when watching an entire sequence from the beginning to end. The individual channels display few actions within themselves and sometimes offer little association between song and video, resulting in a comparatively disjunctive viewing experience.

The slow editing pace applied in these individual sequences may simply be attributed to the enormous effort required to create these video channels. It is therefore likely that this editing pace was used, on the one hand, as a reasonable compromise concerning the overall quality of the music video. Viewers are required to switch between these channels and thereby individually influence the notion of pace. In fact, this slower pace and limited amount of actions and events in individual channels may even motivate the viewer to switch channels. On the other hand, this slower pace may also be useful in order to improve the overall appearance of the sequence when viewers switch channels quite frequently.

Like Chris Milk, Vania Heymann also referred to the use of technology, as well as technological metaphor, to create an engaging viewing experience.⁹⁶ The notion of realism in the images that are particularly familiar to American audiences results in a surreal and entertaining effect that could not be achieved otherwise. Heymann furthermore transformed the "passive act" of "channel-flipping"⁹⁷ into an active and engaging experience.

95 Edwards, Gavin. "Inside Bob Dylan's Brilliant 'Like a Rolling Stone' Video", *Rolling Stone*, 20 November 2013, <http://www.rollingstone.com/music/news/inside-bob-dylans-brilliant-like-a-rolling-stone-video-20131120>.

96 ibid.

97 ibid.

Heymann stated that these creative methods also served to convey social commentary: “I’m using the medium of television to look right back at us – you’re flipping yourself to death with switching channels [in real life].”⁹⁸

In several ways, *Like a Rolling Stone* closely adheres to Vernalis (2013) eight characteristics for online video aesthetics, summarised in Chapter 5.1. Pulse and reiteration are created through user engagement, with the activity of switching channels and potential repeated viewing to discover more of the video. Irreality and weightlessness are grounded within the images’ parodic and fleeting nature. Their variability depends on the video’s replication of media mechanisms.

A recurring motif throughout many of these channels is a focus on the streets of New York. It appears that several sequences were shot in New York, and Dylan’s lyrics mentioning life “out on the street” are often illustrated with associative images.

The different sequences are largely unrelated to the song and its lyrics, but occasionally moments of synchronisation are used when both song and video allowed for such an opportunity. All these video channels only seem to relate to each other in the viewer’s mind. As Dancyger (2011, 388) explained, “[o]ne of the remarkable elements of editing is that the juxtaposition of any grouping of shots implies meaning”. This is emphasised by the fact that a part of the editing is left to the viewer. The author’s control in causing perceptive connections in the viewer is based on the amount and type of audio-visual content provided. Being a very well known song, the music video can build on the viewers’ familiarity and generate humour by letting various protagonist’s exclaim the song’s words in all manner of situations. Additionally, the cynicism and contempt expressed by Dylan was combined with images in such a way that they create an indirect commentary on popular media.

While most of the videos created with Treehouse tend to provide interactive narratives that let viewers either explore the narrative space or change the course of a given sequence, *Like a Rolling Stone* offers a purely non-linear experience with no logical beginning and end or otherwise structuring element, except for its consistent synchronisation with Dylan’s lyrics. More important in this video, Yoni Bloch explained, is the effect of variation, and perhaps

98 Hernandez, Brian Anthony. “Bob Dylan’s ‘Like a Rolling Stone’ Interactive Video Mimics TV Surfing”, Mashable, 20 November, 2013, <http://mashable.com/2013/11/19/bob-dylan-like-a-rolling-stone-interactive-video/>.

what happens in a viewer's mind by associating these video channels in different ways every time they watch it. According to Bloch, "you'll always miss something because you can't watch everything at the same time".⁹⁹

The concept used in this interactive music video was also applied in the video for Aloe Blacc's single *Love is the Answer* and will likely be repeated in the future because, technical challenges and increased production time and effort aside, it is a simple and effective way to provide variation and entertainment value. Viewers don't even necessarily need to engage with these videos, as opposed to *The Wilderness Downtown*, which relies on user input and requires active viewing. Future implementations of the composition method used in *Like a Rolling Stone* and *Love is the Answer* might utilise random switching between available sequences when the viewer is not interacting with the music video, thereby automatically creating variation.

While these videos are often regarded and marketed as curiosities and examples for innovative use of online video, the type of interactivity and non-linearity applied in Interlude's videos may become increasingly modular with regard to video creation and composition methods, and might also provide several advantages for implementation of a specific platform or aggregate that showcases a larger number of these videos.

Among the greatest challenges for these types of non-linear videos is the amount of video production involved, even with the technical realisation for interactivity simplified by a software tool such as Treehouse. Therefore, the application of these methods may rely either on highly experimental works that are created primarily for the purpose of showcasing technology and innovation, or for high-budget projects.

5.3.2.2 *Interactivity*

While the methods of interactivity are comparatively simple from the user's point of view, they have tremendous impact on the entire presentation of the music video. The viewer can only choose to go up and down the channel list, i.e., the viewer cannot influence the order of sequences. The music video director therefore has necessarily given up a large amount of control, not just by letting viewers switch channels, but also by having the video start on a

99 ibid.

random channel. As an alternate method of control, the reduced amount of events and the lack of narrative complexity within each individual scene may stimulate viewers to switch channels. This interactive viewing experience can be seen to take precedence over the viewing of a precisely edited sequence.

This shows that, although some of the author's activities are passed on to both the computational system and the viewer, the author still exercises control in other ways. In this case the images themselves may cause the viewer to perform certain actions. The control over the sequence therefore becomes less tangible and more prescriptive and suggestive instead.

Murch (2001) suggested that editing is a technique which has to predict as well as guide a viewer's eye. In *Like a Rolling Stone*, the author may be less concerned with the viewer's gaze, but rather with his or her motivations and subsequent actions.

5.3.2.3 Structure

Like a Rolling Stone is a prime example for non-linear video, and therefore any description of the video's structure will necessarily be vastly different from characterisations of linear music videos. While *Like a Rolling Stone* consists of sixteen completely individual video sequences, the editing within these sequences plays a minor role compared to the overall context of the music video, aside from the fact that they are typically edited to (1) best represent the respective television trope they are meant to embody and (2) function as seamlessly as possible when the viewer is switching channels.

In the following, each individual channel will be briefly discussed with regard to its overall content, editing dynamics and audio-visual connections.

Romantic Comedy	A couple is seen walking out of a New York residence while their conversation follows the song's lyrics. After the first half of the video, the couple is seen walking down the street, and during the last third of the video their conversation continues in a café. The visual rhythm is barely synchronised to the music but is driven by the occurrence of vocals and the sequence's embedded narrative, which largely adheres to conventions for visual continuity and to Bordwell & Thompson's (2010) definition of classical Hollywood cinema.
Infomercial	The infomercial features a female and a male host who are advertising a vacuum cleaner. While this sequence strongly follows common aesthetics associated with

	<p>infomercials, the overall pace of the song and vocals make the progression comparatively slower than a viewer may be used to. In the absence of vocals, this sequence even tends to depart from the aesthetic convention, and its editing then stronger resembles the discontinuous editing of music video than might be the case in its real-life equivalent. After the first half of the video, it becomes apparent why a vacuum cleaner was chosen as the advertised product, when Dylan is heard singing: "as you stare in the vacuum of his eyes". This moment in the song is emphasised by a close-up of the female host mouthing the words. Similarly, the line "How does it feel" is highlighted especially by the male host when he handles the vacuum cleaner stroking pieces of carpet used to showcase the product's effectiveness.</p>
Music Video	This channel shows the rapper Danny Brown lip-synching to the song in a representation of an official music video for the musician. Here, Danny Brown is seen on the street, predominantly eating fast food, which seems to relate to Dylan's lyrics about "having to be scrounging your next meal" and "nobody's ever taught you how to live out on the street". The overall editing pace appears to be slower than one would imagine in a real music video. Some of the shots in the sequence seem unnaturally long to the extent that they may become either humorous or dull.
Performance	A real live performance of the song is shown. The overall progression and aesthetics of this sequence are typical for this type of performance video.
News	The sequence intercuts between a male host reporting the news and several news segments, which seems to be a rather unnatural and increasingly musical treatment for the editing of a news report. Naturally, a viewer would be accustomed to news shows being tightly organised and edited in order to fill the program's allocated time as effectively as possible with audio-visual content. However, due to the lip-synching to Dylan's vocals, the host's speech appears to be unnaturally slow. The sequence makes use of the aforementioned common motif of New York City streets when the lyric comes up in the song by showing the stabbing of a person involved in a drug crime. Several shots in this sequence representing CCTV footage again appear unusually long, as they are used to depict lip-synching of the vocals. Eventually the sequence also features a weather report with similarly odd pacing.
The Bachelor's Roses	Dylan's lyrics are used here to drive the progression of interviews and commentary by the different protagonists as well as a conversation between a group of women in a living room. In this channel, the overall editing pace appears more authentic than in other sequences due to the nature of the footage and the frequent cutting between the protagonists. Song and video seem to connect in what appears to be a fight between two of the female protagonists as one stops and mouths the words to the line "you've gone to the finest schools, alright Miss Lonely". However, this synchronisation is not consistent as the following lines are divided between various protagonist's in the room. Perhaps the association simply arose by chance since the visual performance had to be paired with the song, or perhaps the director intentionally ensured that no such close synchronisation would persist anywhere in the video, in order to maintain its ambiguity. The last third of the video focuses on the male protagonist showing him

	<p>spending time with one of the women. Again, while viewing this section, it is easy to fall into the temptation of associating the lyrics about precious gifts and diamond rings with what is depicted in the sequence, but no close and obvious relationship between these lyrics and the images is provided.</p>
Documentary	<p>A documentary about New York City, aptly titled "The Streets of New York" is shown, featuring interviews with alleged experts and vintage footage of the city. The sequence's pace and progression largely appears authentic. However, lip-synching of the song is again prioritised, and therefore a large fraction of the sequence consists of shots depicting the interviewees. While the entire sequence is an example for the music video's motif described above, a closer synchronisation between song and visuals again appears when the lyrics refer to life on the streets, with a shot showing a man selling newspapers on a street, followed by several shots of New York City traffic. A further connection is made when Dylan asks "Do you want to make a deal?" and the video shows the words "make a deal" as an advertisement on Times Square.</p>
Business News	<p>A male news anchor is seen interviewing a female reporter on a street, presumably in New York as well. Again, the overall pacing of the sequence and representation of the protagonists' conversation appears humorous, especially when the two reporters keep asking each other "How does it feel?" followed by long pauses of the two seemingly staring at each other silently. Similar to the reality TV show, it appears as if the two reporters were having an aggressive conversation in parts of the sequence.</p>
Cooking Show	<p>This sequence shows a seemingly disgruntled chef preparing a dessert. Since the focus is directed towards the chef's activities and the viewer is made to follow the preparation of the recipe, the overall editing pace appears comparatively natural.</p>
Pawn Stars	<p>This sequence is an exact imitation of the original American reality television series <i>Pawn Stars</i>. The show's protagonists are shown discussing the value of a vintage guitar in varying shots depicting conversations and individual interviews. The overall pace of the sequence is probably much slower than the pace of the original it is based on, especially due to Dylan's frequent pauses between individual lyrical lines. Once again, as Dylan mentions the street, the protagonist Michael is seen being interviewed on a street. A further connection is made when one of the salesmen, who has been discussing the guitar's value with Michael, the guitar's owner, asks, "Do you want to make a deal?" and later advises "You better pawn it, babe".</p>
Fashion Show	<p>This sequence intercuts between a street interview, the reporter in a studio and shots of a runway show. The overall progression is disjunctive and discontinuous, probably to imitate the look and feel of particular fashion shows. As a whole, the sequence appears comparatively random and, like many of the other sequences, slow-paced. After the first half of the video, the sequence cuts to a different segment depicting a trailer for a television series. The sequence then cuts back to the initial show which continues to progress as described above.</p>
The Price Is Right	<p>This is another exact imitation of a real American television series, in this case the game show <i>The Price is Right</i>, hosted by Drew Carey. The overall depiction of the show appears comparatively authentic compared to the original, effectively using the</p>

	colourful studio and various actions performed by the candidates to progress the sequence efficiently until the last third of the sequence. After the candidate has won, Carey is shown addressing the audience until the candidate is seen again towards the end.
Maron	<i>Maron</i> is an American comedy television series, which has also been imitated for a sequence in this music video. Here, Marc Maron is shown conducting a radio interview. In the context of a humorous scene from a television show, the overall editing and progression of the sequence seems authentic and believable. Lip-synching of the vocals is used effectively to generate the conversation with typical dialogue editing.
Tennis Match	This sequence shows a tennis match, which is predominantly illustrated via close-ups of the players, as well as members of the audience, lip-synching the song. A sports reporter in a studio further commentates the game via Dylan's vocals. Again, the editing seems somewhat unnatural when the sequence is viewed in its entirety, as it frequently cuts between the reporter, close-ups of the players and occasional moments of the actual match, which, similarly to some of the other news report-type sequences, uses a more musical type of editing characteristic of music videos.
Property Brothers	Here, the Canadian reality television show <i>Property Brothers</i> is imitated. The sequence depicts the show's two protagonists inspecting a property with a couple who is assumingly willing to buy it. An association between song and video can be drawn when Dylan sings, "How does it feel to be without a home?" However, this connection may be arbitrary, since the conversation between the four rather seems to be revolving around several details of the building. The editing of the sequence appears comparatively authentic, as the building, cutaway animations and conversations between the protagonists with sometimes fast camera movements are used to construct the simple narrative of this show.
Zoey and Socks	This sequence shows an animated children's television series. The overall progression revolves around the two characters presumably teaching the spelling of arbitrary words. The sequence's pace seems appropriate and natural for this type of program.

Table 5.3.1: Like a Rolling Stones Music Video Channels.

5.3.3 Do Not Touch

*Do Not Touch*¹⁰⁰ is an interactive music video created for the band Light Light and their single *Kilo*, released in April 2013. The video consists of various animated and live-action footage, as well as a number of recorded cursors guided by a succession of instructions displayed to the viewers (cp. Figure 5.3.3). In these instructions, viewers are asked to perform certain actions with their cursor on top of the moving images. Their cursor movements are recorded and gradually added to the video for every following viewer to see. This turns the video into a

100 Do Not Touch. <http://donottouch.org/>.

crowd-sourced and ever-changing viewing experience. With this basic concept, the video offers a playful, game-like experience for viewers.

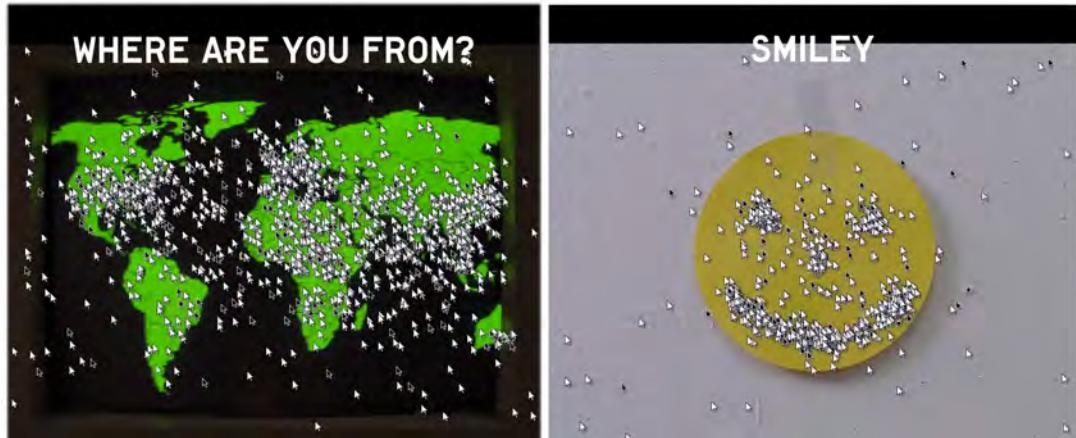


Figure 5.3.3: Screenshots of the *Do Not Touch* interactive video (*Do Not Touch*, <http://donottouch.org/>).

According to the authors, the video for *Do Not Touch* is based on the assumption that, with evolving technologies and the prevalence of touch screens, the cursor, used for interacting with digital software and media, may disappear:

*"After 50 years of pointing and clicking, we are celebrating the nearing end of the computer cursor with a music video where all our cursors can be seen together for one last time."*¹⁰¹

Furthermore, the authors explain that *Do Not Touch* is a highly experimental project that is based on the exploration of a variety of questions:

*"As always, we do not know what will happen with this project. Will people care? Will they follow our instructions? Will patterns emerge which we had not predicted? How many cursors will be too many? A thousand? Ten thousand? Fifty thousand?"*¹⁰²

According to Jonathan Puckey, co-founder of Studio Moniker who produced *Do Not Touch*, the video received approximately one million views within its first two days after release.¹⁰³ At

101 ibid.

102 Visnjic, Filip. "Do Not Touch - Celebrate the nearing end of the humble computer cursor", *Creative Applications*, 15 April 2013, <http://www.creativeapplications.net/featured/do-not-touch/>.

103 Matulef, Jeffrey. "Crowdsourced interactive music video Do Not Touch is a fascinating sociological experiment", *Eurogamer.net*, 19 April 2013, <http://www.eurogamer.net/articles/2013-04-16-crowdsourced-interactive-music-video-do-not-touch-is-a-fascinating-sociological-experiment>.

the time of this writing, over 4.3 million cursors have been recorded.¹⁰⁴ For this reason, only between 800 and 1000 cursors are shown any time the video is viewed.¹⁰⁵

Media coverage of the video again demonstrates the dominance of the video and its innovative creation over the song around which the video is meant to revolve. The title of one article reads “Moniker’s latest crowd-sourced music video gets your mouse warmed up”, which illustrates the focus on the design studio, and therefore the importance of the video itself.¹⁰⁶ Another article mentions the song in only one sentence after several paragraphs: “And oh yeah, the music video is by an Amsterdam based band called Light, Light, and their song’s called ‘Kilo’, instead of ‘Do Not Touch’”.¹⁰⁷

5.3.3.1 Content and Editing

The video begins by informing the viewer that their cursor will be recorded throughout the video's entire duration. After this, simple instructions are displayed to guide and misguide, for those viewers who like to go against the tide, cursor movements.

The events first take place on a computer screen depicting monochrome graphics that viewers are asked to interact with. These interactions range from simple questions that can be answered by pointing to certain parts of the displayed graphics, to a game-like interface that requires viewers to fulfil certain tasks by moving their cursor along particular paths. Eventually the video zooms out from the computer screen and reveals a studio where the band is performing the song. The instructions given to the viewer become more varied and open-ended. Viewers are asked to form the opponent of a boxer, create a smiley face on a blank round shape and mimic to play several instruments.

Interactivity and variation are not created by direct interaction that triggers immediate feedback to the user via a change in visual appearance, but simply by recording the user's input, in this case cursor movements, and eventually adding this data to the video. With the

104 Eve Dullart, e-mail message to author, 4 December, 2014.

105 ibid.

106 Bec, Alex. “Moniker’s latest crowd-sourced music video gets your mouse warmed up”, *It’s Nice That*, 15 April 2013, <http://www.itsnicethat.com/articles/monikers-latest-crowd-sourced-music-video-gets-your-mouse-warmed-up-for-the-week-dot-dot-dot>.

107 2pause. “Do Not Touch”, <http://www.2pause.com/do-not-touch/>.

successive addition of more and more of these cursors, an unusual kind of narrative is created that emerges from the participant's collective behaviour over time. Thus, a comparatively unpredictable visual layer is added to the video that essentially constitutes its primary aesthetic and narrative value.

In this context, the recording of cursor movements becomes both a constructive and a destructive element for the music video's progression. Previously recorded cursors tell new viewers what has happened before, influencing their behaviour. However, the video also conveys an inevitable destruction with increasing addition of individual cursors which obscure the initial images that served as guides for their behaviour.

The notion of participation is so strong in this video, that some articles do not speak of watching it, but participating in it.¹⁰⁸

Do Not Touch is therefore a linear, but non-narrative video that foregrounds the viewers' playful interactions. According to Roel Woulters, fellow co-founder of Moniker, “[p]laying along with your cursor requires a lot of attention, which can move the ‘story’ somewhat to the background”.¹⁰⁹

5.3.3.2 *Interactivity*

Through the use of crowdsourcing, *Do Not Touch* does not directly give control to the individual user, but to the video's audience at large. The individual user will not necessarily be primarily motivated to effect change on the video, but rather to participate in the video's development and engage in the time displaced communal experience.

The creators of *Do Not Touch* explicitly emphasised that they were not interested in exercising control over the video, but instead wanted to see how users will react to and interact with the video. The production therefore aimed to provide users with a space where they can playfully engage with the images. In viewing the video, the foreground is formed by the users' cursors and remains the primary focus of attention.

108 McCarty, Brad. “Do Not Touch: A crowdsourced Music Video, Where Your Mouse Pointer is the Star”, *The Next Web*, 28 April 2013, <http://thenextweb.com/shareables/2013/04/28/do-not-touch-video/>.

109 2pause. “Do Not Touch”, <http://www.2pause.com/do-not-touch/>.

While this option does not seem to have been publicly exploited by Studio Moniker, the large amount of data generated by users all over the world provides the potential of restructuring the cursors according to different time zones, countries or other parameters, thereby adding another dimension to the virtual narrative created by the audience.

With this video, the author does not give up control over the progression of the sequence. However, with the visual and interactive focus clearly lying on the additional layer formed by the audience's cursors, a high level of unpredictability is introduced. Although this unpredictability was intended in this case, this concept may not be transferable to other projects, since many authors may not want to grant this kind of control to the audience because many viewers will naturally act against any given instructions, resulting in visual outcomes that may not be desirable for the author.

5.3.3.3 Structure

Intro	A green dot followed by a world map is shown on the computer screen. Viewers are first asked to stay within the green dot and then to point to the area where they are from on the world map.
Verse 1	The map's colouring is inverted and viewers are now asked to indicate where they would like to go. The next screen displays two pictograms for a male and female figure. Viewers are asked to identify with one of them. Again, the picture is inverted and viewers are asked which one they "prefer".
Verse 2	The screen turns green and reveals a map of different pathways that are moving from right to left on the screen. Viewers are asked to stay "in the green zone".
Chorus 1	The previous scene continues.
Verse 3	The scene reveals the band name "Light Light" on the same screen.
Chorus 2	The previous pathways turn into green blobs that move around the screen in varying shapes. The instruction to stay within the green zone persists.
Bridge	The video zooms out from the screen and reveals the studio in which it is situated. An establishing shot shows the band performing the song and a naked model sitting on the ground. The video cuts to a shot of a boxer and viewers are asked to form her opponent.

- Verse 4 The video cuts to a close-up of one of the singers. Viewers are asked to form a mask on her face. Then a close-up of the bassist's hand and bass neck is shown. Viewers are now asked to "play the bass".
- Chorus 3 A shot of the naked model is shown and viewers are asked not to touch her. Next, a shot of another singer is shown and viewers are again asked to form a mask. This is followed by a shot revealing a piece of paper with a simplified timeline indicating "birth" and "death", and the viewers are asked to point to their relative position on this timeline.
- Chorus 4 After continuing to show the previous scene, the next shot depicts a simple gradient and viewers are asked to "avoid each other", leading to an immediate dispersion of cursors all over the screen. The next shot shows a synthesiser, and viewers are now again asked to "play" the instrument. This scene is followed by a blank round yellow shape with the simple instruction: "Smiley". The final scene shows a small arrow on a white wall and users are asked to "form a point".

5.3.4 Discussion

The case studies in this section were conducted in order to gain a better understanding of interactive music video, the videos' methods of interactivity, how they differ from linear music videos, and how control over the presentation of a sequence is transformed and maintained from an author's point of view.

The three interactive music videos discussed for this purpose are based on three different concepts defining the videos' progression, content, interaction and user engagement. These concepts consist of

- 1) a multi-window approach for music video presentation, featuring customisation,
- 2) a simulation of the television viewing experience which allows for channel switching to provide interactivity, and
- 3) a sequence that features a gradual development of graphics with the use of crowdsourcing.

These case studies suggest that music videos which make use of interactivity and non-linear methods for video sequencing typically reduce the level of narrative logic in order to foreground their experimental and entertaining value.

While *The Wilderness Downtown* shows a gradual progression, guiding the viewer's feeling state throughout the video and giving the impression of a vague narrative, it does not feature a discernible story characterised by a sequence of discrete events. In comparison, both *Like a Rolling Stone* and *Do Not Touch* can be considered as entirely non-narrative. Their potential narrative value lies in the communication of ideas, the role of mass media in our lives and the impending disappearance of the cursor. *Like a Rolling Stone* therefore simulates the experience of watching television in a simplified set-up and completely gives up control over what is seen at any given point during the duration of the music video. In comparison, *Do Not Touch* does provide a completely linear and non-variable video structure, but presents the viewer with a series of scenes as templates for interaction.

Four different types of interactivity were used in these music videos:

- 1) preliminary customisation,
- 2) the use of mouse movement and key strokes to influence the content of a scene,
- 3) channel switching and
- 4) gradual modification of a sequence through crowdsourcing.

These types of interactivity lead to different modes of user engagement. Customisation is used to personalise the viewing experience. By creating possibilities to exercise direct influence on the images themselves via mouse movements and keystrokes, the user is prompted to explore and discover the interactive features of a given video. Channel switching provides another mode of exploration by giving the user the chance to see something different every time he or she watches the music video. Crowdsourcing and therefore the gradual modification of a music video, as demonstrated in *Do Not Touch*, allows the user to participate in a gamified social experience.

It has been established that music video is frequently seen as non-narrative or at least ambiguous in this regard. The three interactive music videos examined in this section suggest that, with increasing engagement and interactivity, the video's narrative diminishes, either to make room for playful interaction, or to encourage extrapolation and emotional engagement.

From the author's point of view, we have to look at a new mode of treating video material, which is based on a changed purpose of a video sequence with regard to viewer engagement,

and the way in which the author exercises control over the sequence outcome adopts different functions. Any time the author allows a computational system or a viewer to influence a video sequence, he or she is giving up a certain degree of control, resulting in a certain level of unpredictability. The importance of precision in sequence composition is relativised in order to make room for these different modes of viewer engagement and sequence variability. The more a sequence's progression and visual appearance can be altered, the more an author will have to consider other methods of maintaining a certain level of control to ensure an adequate coherence toward the expected outcome and effect.

5.4 Music Video Categorisation

In the first decade after the introduction of MTV and the definitive design of the music video, much research was undertaken to better understand music video and its context. Several approaches to define music video were based on isolating typical themes, categorisations and genre specifications (Gow 1992). The most common typologies found in literature include and revolve around categories and themes such as performance, concept, narrative and non-narrative clips (Lynch 1984, Goodwin 1992, Gow 1992, Keazor & Wübbena 2007).

A different type of categorization was suggested by Holdstein (1984), who distinguished videos with explicit political themes, videos that build on the U.S. film musical, and fantasy videos. Kaplan (1988) categorised music videos from the mid to late 1980s according to ideology: romantic, socially conscious, nihilist, classical and post-modernist. Jones (1988) focused on narrative differentiation and defined three types of narrative representations according to their level of abstraction: mimetic, analogue and digital. Gow (1992) formulated a typology revolving around what he considered the central element in music video: the music itself, and therefore its visual representation as performance. He established the following categories:

- the anti-performance piece,
- the pseudo-reflexive performance,
- the performance documentary,
- the special effects extravaganza,
- the song and dance number and
- the enhanced performance.

Building on Gow's categories, Railton and Watson (2011) proposed one of the most recent categorisations of music video, also considering the definitions by Lynch (1984) and Jones (1988). They suggest four different types:

- the pseudo-documentary music video,
- the art music video,
- the narrative music video and
- the staged performance music video.

Many categorisations were faced with the criticism that a definition of the music video is nonsensical because a pure representation of a music video category will rarely appear. Goodwin (1992, 71) suggested that music videos are, "if anything, 'typically' rhizomatic". Keazor & Wübbena (2007) argued that classifications, as mentioned above, might be more of an obstacle than a help because they may blur the vision of the true essence of a given music video.

In another attempt to find ways of referring to different types of music video, Vernallis (2004) focused on a scale of narrativity, from highly narrative to non-narrative videos and suggested using modes along that continuum rather than defining a set of categories. She argued that the narrative progression in music videos will often either display a certain degree of classic narrative with a beginning, middle and end or develop through "slow disclosure" (Vernallis 2004, 197). Concerning content, she further suggested to distinguish two typical themes: first, the completion of a single process or goal, and secondly, videos that perform like a series or lists of events that follow each other without a clear narrative progression. A combination of these themes may be, for example, the travelogue or "mechanical chains of causality in which the camera tracks the flow à la Rube Goldberg's machines" (Vernallis 2004, 20). Although the completion of processes can usually be considered to constitute a narrative, such videos will often appear as non-narrative until the end, since a viewer may only become aware of an existing narrative upon seeing its resolution. Another typical structural theme is the extension of a performance space where the settings themselves become a narrative element.

Chapter 5.1 included a discussion on the many reasons and motivations for music video's aesthetic nature. In this case, it is necessary to foreground the music video's structure and its narrative form. When compared to film, music video appears to show more similarity to the television commercial, and some authors argued that music video necessarily had to borrow

aesthetic techniques from advertising (Jhally 1995, Strachan 2006). Jones pointed out that this similarity does not occur because they are both selling a product, but rather because they both operate “within similar social, narrative, economic and technological limits” (Jones 1988, 23). Instead of focusing on narrative clarity, music video seeks to highlight music and musicians, thereby creating its own narrative logic (Vernallis 2004, Peverini 2010). Jones (1988) called this the digital narrative. Chapter 5.1 examined other descriptions of narrative, including definitions by Abbott (2008), who took a rather open stance on this subject, loosely referring to narrative as an organisation of events, Todorov (1971), who gave a more detailed overview concerning structural parameters that must be met in order for something to classify as a narrative, and Branigan (1993), who, in addition to processual structures of narrative, provided an alternative list of methods to organise information in narrative and non-narrative ways.

With the help of the case studies outlined in the previous section of this chapter that examined both linear and interactive music videos, an attempt was made to better understand the influence of online media on current music video aesthetics. The analysis showed that, whereas linear music videos may have undergone a limited transformation in order to adapt to a new digital medium, interactive music videos constitute, in many ways, an entirely new direction for the form in general, and therefore introduce new concepts for audio-visual composition. While interactive music video seems to establish a field of its own, examples are still highly experimental and comparatively rare, and the coming years will show in which directions these videos are going to develop. At the time of this writing, it is unlikely that definitive overarching patterns and common conventions in interactive videos can be identified that will help in the further investigation and creation of non-linear music video editing methods.

However, the results from the previous case studies on non-linear videos helped to extend and confirm several theories of music video aesthetics. Yet, in order to support the design of an algorithmic music video editing system, further details concerning generalisable patterns of structure and aesthetics are required. The following analysis will therefore investigate a larger body of music videos in order to uncover possible trends and music video categories to better understand and synthesise its underlying logic and rhetoric structure.

In this analysis of 88 music videos taken from the list of *Triple J's Hottest 100 2010*,¹¹⁰ three types of classifiers were defined based on literature and the case study results, and tested for their relevance for an application in the context of categorisation, editing conventions and potential definition of editing patterns. These classifiers are music genre, content types and levels of narrative.

The Hottest 100 is an annual music poll conducted by the Australian radio station Triple J, calling it “the world’s largest music democracy”.¹¹¹ In the most recent list of *Hottest 100*, for instance, voters from seven countries participated (43% in Australia) with 1.49 million votes cast.¹¹² *The Hottest 100* polls tend to favour popular music genres such as rock and indie, and, unsurprisingly, feature a comparatively large amount of Australian music (cp. Appendix A).

5.4.1 Music Video Editing Characteristics

In order to analyse this comparatively large body of music videos, it was necessary to find efficient descriptions to summarise the music video characteristics outlined throughout this and the previous sections. In addition, it was necessary to find descriptors that focus predominantly on the technical process of editing, largely ignoring the visual content within shots.

The analysis process involved repeated observation of each video on the list during which features were coded and their occurrence counted. Next, the data and their descriptors were filtered and summarised into a small number of parameters that were deemed relevant for the proposed technical experiments. The selected parameters are the videos’ linearity, orientation to the song’s form, rhythmic editing, segmentation, lip-synching, lyric references and shots per minute.

These characteristics were chosen in order to examine the variability of sequences with regard to the order of shots and its impact on the meaning of the overall video. In this context, the music videos were also examined for possible segmentation in order to reveal to what extent events correlate with musical sections. A further point of investigation was the observable

110 Triple J. “Triple J’s Hottest 100”, ABC, <http://www.abc.net.au/triplej/hottest100/>.

111 Triple J. “Hottest 100 Archive”, ABC,
<http://www.abc.net.au/triplej/hottest100/archive/about.htm>.

112 Triple J. “Triple J’s Hottest 100 2013”, ABC, <http://www.abc.net.au/triplej/hottest100/13/>.

musicality in the sequencing of shots. For this purpose, an average value for the video's editing pace based on the number of shots per minute was also documented. Finally, the frequency and extent of audio-visual synchronisation was measured with regard to performance and other actions that relate to the sonic elements of a song.

5.4.1.1 *Linearity*

Here, linearity refers to the video's structural rigidity and constitutes an essential element of narrative clarity. Linearity was measured in order to examine these videos' variability (cp. Chapter 1) and their adherence to previously outlined characterisations, for instance, with regard to Jones' (1988) definition of the digital narrative and Dancyger's (2011) description of MTV style.

This measure therefore indicates the potential for changing the existing video structure without significantly changing its meaning. Videos with a strictly narrative structure and a low level of variability were classified as linear; those not following such a structure and possessing a high level of variability as non-linear.

5.4.1.2 *Orientation to the Song's Form*

Music video scholars from the 1990s onwards have stressed the importance of addressing music video in terms of the song that it represents, which, in the first years of MTV during the 1980s, was often neglected (Björnberg 1994, Vernallis 2004, Railton and Watson 2011). Strachan (2006, 189), for instance, emphasised that “the tropes and techniques common in the music video format are better understood in the relation to the syntactical characteristics of popular music”, and that the editing therefore attempts to “visually echo and reflect the rhythmic, structural, and melodic properties of a given musical text”.

The degree to which videos followed a song's form was measured by how often the video clips displayed visual and narrative changes with respect to the musical segmentation. This means, for example, that visual differences align with chorus, verse, and bridge structure.

All the videos analysed in Chapter 5.2 showed some degree of orientation to musical segmentation but did not strictly follow this convention throughout the entire duration of the

music video. The following analysis will show in more detail to what extent this pattern occurs across different types of music video (and what these types are).

5.4.1.3 Segmentation

According to Vernallis (2004, 163), music videos can often be observed to consist of a number of visually distinct segments because “a video will often draw attention to a song’s sectional divisions in order to help the viewer remember the song and its structure”. The segmentation classifier is therefore closely related to the measure of orientation towards the song’s form, which will become more evident in the following discussion.

The differences between segments can be either dramatic or subtle. With a focus on technical implementation, a segment is defined as a series of shots that show a clear distinction from the previous and following segments. A segment may have a number of shots with additional visual characteristics not shown in previous segments (cp. Figure 5.4.1) or it may be distinctly different from other segments as a whole (cp. Figure 5.4.2).



Figure 5.4.1: Schematic illustration of the additive segmentation of video content.

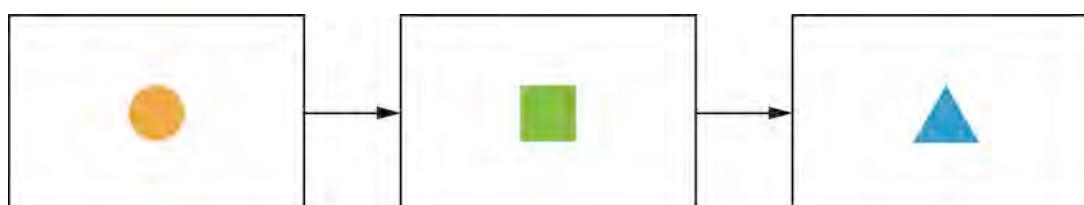


Figure 5.4.2: Schematic illustration of the substitutive segmentation of video content.

By comparison, those video clips classified as having few or no segments show a less disruptive, more continuous flow of images with either an abstract setting or a strong narrative, referred to by Vernallis (2004, 197) as “slow disclosure”.

5.4.1.4 Rhythmic Editing

This characteristic provides a measure on how often the shots were synchronised with the musical beat and metric boundaries. This measurement was based on the assumptions that a music video's visual rhythm is "closely connected to the song" (Vernallis 2004, 27).

Confirming some of Vernallis (2004) theses concerning the visual highlighting of hooks and choruses, the case studies in Chapter 5.2 suggested that beat driven songs, e.g., *Gangnam Style* and *Call Me Maybe*, which are well known for being "catchy", will often use editing techniques that highlight such notable moments in a song.

The analysis of this characteristic is based on simple and obvious features: strong drum beats, vocal highlights and melodic changes, among others. Björnberg (1994) suggested a more in-depth approach to the analysis of the music in music videos, for instance based on characteristics such as discursive repetition, demarcation and symmetry. However, for the purpose of creating an overview of current music video trends and their possible implementation in a generative system, less detailed descriptions were used.

5.4.1.5 Lip-synching

Lip-synching is easily measured and occurs when the musician's performance is in shot and in time with the music. Given that most linear music videos will revolve around the musicians and their performance, most videos include at least one short section of lip-synching by the artist or fictional character avatars in the video.

5.4.1.6 Lyric References

Visual references to lyric content include obvious visual depictions of the words being sung, for example with gestures or written words. Occasional appearances of visual references to lyric content were classified as "medium" and regular occurrence throughout the song as "strong".

5.4.1.7 Shots per Minute

While editing pace and rhythm can change significantly during the course of a music video, the shots-per-minute value served to provide an overview of the average editing pace across all music videos. This was done to indicate if certain types of music videos generally require slower or faster paced editing.

5.4.2 Music Genre

So far, literature has suggested that music video has to revolve around the song it represents. Music genre is generally used to classify musicians and their sound within particular sections of popular culture, largely for promotional purposes (Strachan 2006, Railton and Watson 2011). It therefore seems reasonable to associate particular visual features and dynamics with this method of identifying and differentiating popular music.

Railton and Watson (2011, 45), however, warned against using music genre as a classifier for music video categorisation, arguing that “genres of music neither map onto genres of music video nor ... govern the look of any given video”. Chion (1994) compared music video to silent film, and other theorists have mentioned music video’s seeming relation to montage theory, which operated within silent film as well (Vernallis 2004, Gabrielli 2010). This suggests that the video, as part of music video, may operate within its own context as a visual layer parallel to the music.

Music genre was still included in this analysis to test these hypotheses and examine if any of the technical parameters outlined in Chapter 5.4.1 generated observable patterns for this classifier.

Four genres were chosen for this analysis, according to their occurrence on the list of music videos: electronic, hip-hop, indie and rock.

5.4.2.1 Results

In summary, although some differences across genres were observed, the results by music genre are largely homogeneous, at least for the purpose of algorithmic implementation, where distinct values are needed in order to provide a reference system for categorical

differentiation. Perhaps the most distinct genre in this section is electronic. This may be explained by the genre's generally more simplistic representation of performance, which will often lead to a visual style that is orientated towards dance and narrative. This was confirmed by Gabrielli (2010, 91), who observed that "this music genre does not necessarily have the main aim to present the performing artist to the audience".

Some differentiation was also found in the hip-hop genre, which showed a tendency towards either highly linear or non-linear videos and a similar trend in editing pace, with either slow or fast-paced videos. Hip-hop videos also showed the highest percentage of videos featuring visual references to lyrics.

Videos in the rock genre tended to be semi- to non-linear.

In conclusion, knowledge of the music genre alone does not provide sufficient information to distinguish music videos. The data suggests that videos belonging to a particular music genre do not necessarily share a unique approach to visual features. However, some genre-specific tendencies were observed that will be discussed in the following sections. A detailed list of the data discussed in this chapter can be found in Appendix A.

Electronic

Videos belonging to the electronic genre do not display strong trends in linearity, rhythmic editing and segmentation. They tend to show a balanced distribution in visual adherence to the song's form. The genre further exhibits the lowest percentage of weak adherence and the largest percentage of music videos that strongly follow the song's sectional division (cp. Figure 5.4.3). A similar trend was observed for segmentation (cp. Figure 5.4.4). The majority of videos feature lip-synching but have few visual references to lyrical content. However, the electronic genre also features the largest number of videos without lip-synching (cp. Figure 5.4.5). The average editing pace is low to medium.

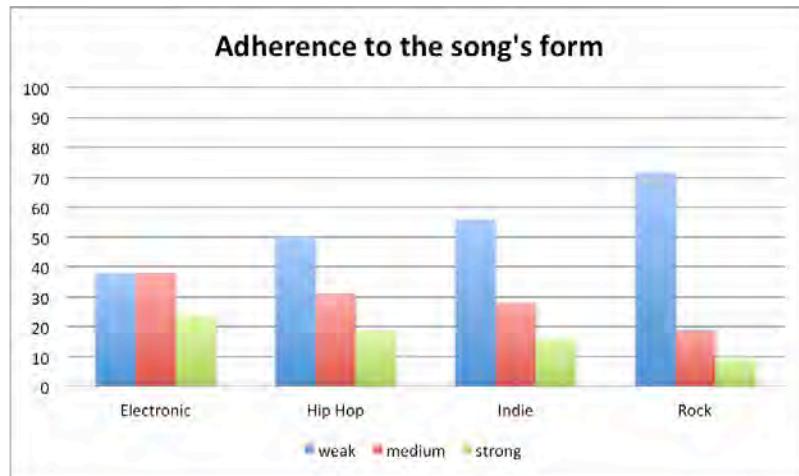


Figure 5.4.3: Adherence to the song's form by genre.

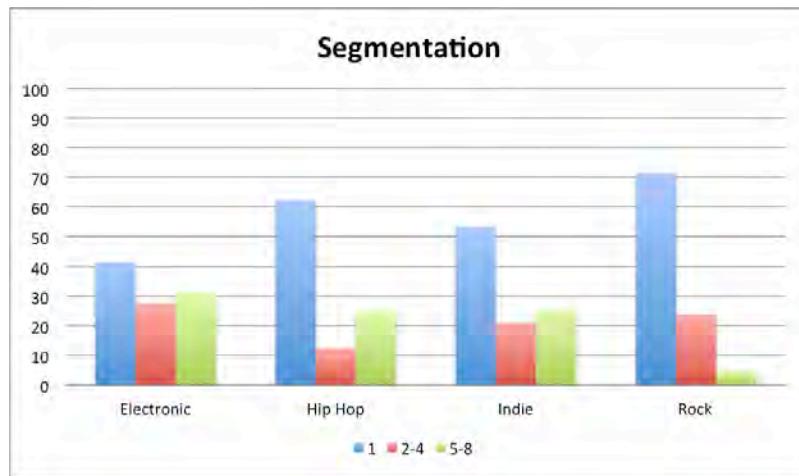


Figure 5.4.4: Segmentation by genre.

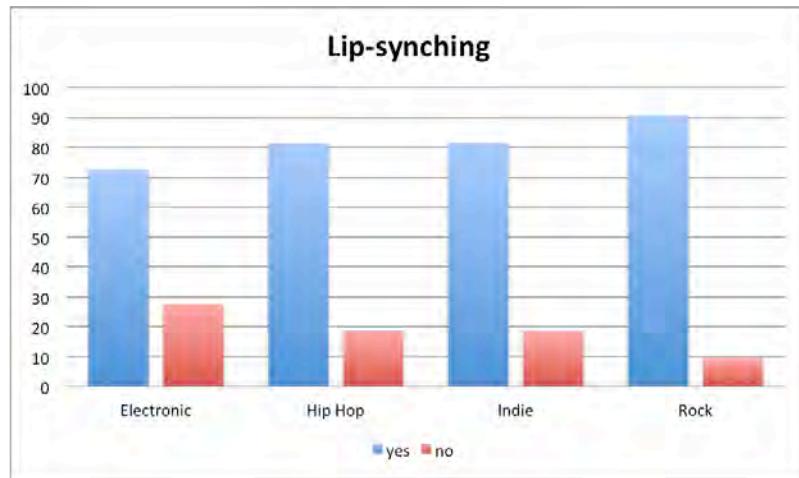


Figure 5.4.5: Lip-synching by genre.

Hip-hop

The hip-hop genre displays no strong trends in rhythmic editing and visual references to lyrical content. Whereas other genres show relatively homogeneous results in linearity, the hip-hop videos in this analysis showed a trend towards either extreme of the linear scale, with the lowest percentage of semi-linear videos (cp. Figure 5.4.6). A similar trend was also observed concerning the genre's average editing pace, and the amount of visual references to the song's lyrics (cp. Figure 5.4.7). The latter characteristic provides for some differentiation between hip-hop and other genres, with hip-hop featuring the highest percentage of lyrical references. A low to medium percentage of videos adhere to the song's form. A large number of videos are slow disclosing or show a structure of 5-8 segments, again showing trends towards either extreme. A majority of videos feature lip-synching. The average editing pace is low to medium. Similar to the electronic genre, templates for this genre may vary greatly.

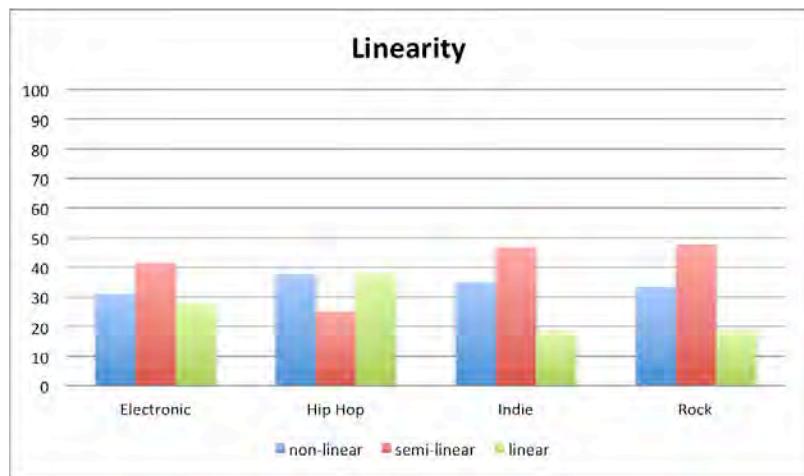


Figure 5.4.6: Linearity by genre.

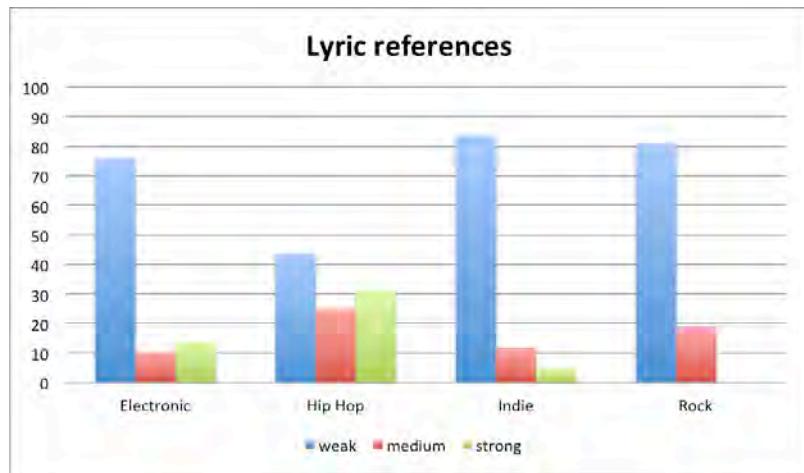


Figure 5.4.7: Lyric references by genre.

Indie

Most of the videos belonging to this genre are non-linear to semi-linear with a weak to medium adherence to the song's form. The indie genre displays no strong trends in rhythmic editing and segmentation. A majority of videos feature lip-synching, but few visual references to lyrical content. The average editing pace is low to medium. Videos in this genre may require less synchronising than electronic and hip-hop, but cuts may appear more frequently.

Rock

Most videos falling into the rock category are non-linear to semi-linear. The videos seem to follow the song's form to a lesser extent than videos belonging to other genres and dominantly show a medium tendency towards rhythmic editing (cp. Figure 5.4.8). This result may be explained by Vernalis' (2004) observation of music videos often showing close synchronisation during the beginning and middle sections of the video, while audio-visual connection and synchronisation become more tenuous towards the end. A similar trend towards a medium value was also observed for the genre's average editing pace (cp. Figure 5.4.9). Most of these videos are not segmented and feature lip-synching, with a few that include visual references to lyrical content. The average editing pace is medium. Rock videos are performance driven and feature a high degree of synchronisation, but generally offer a lot of variability from a narrative point of view.

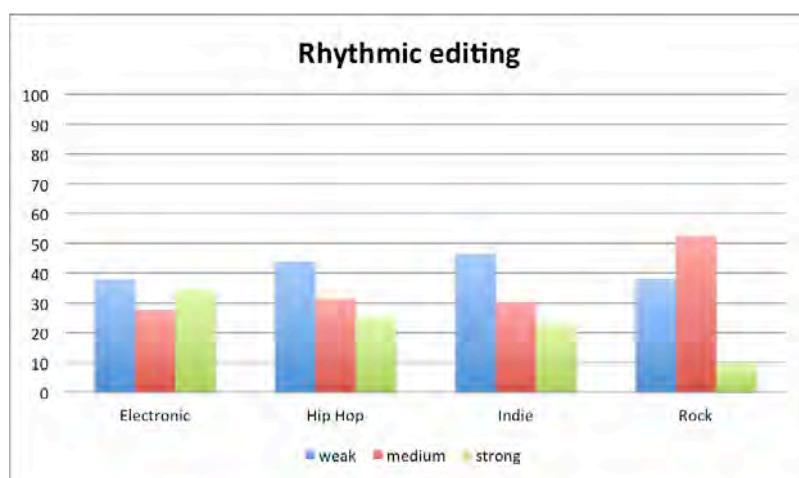


Figure 5.4.8: Rhythmic editing by genre.

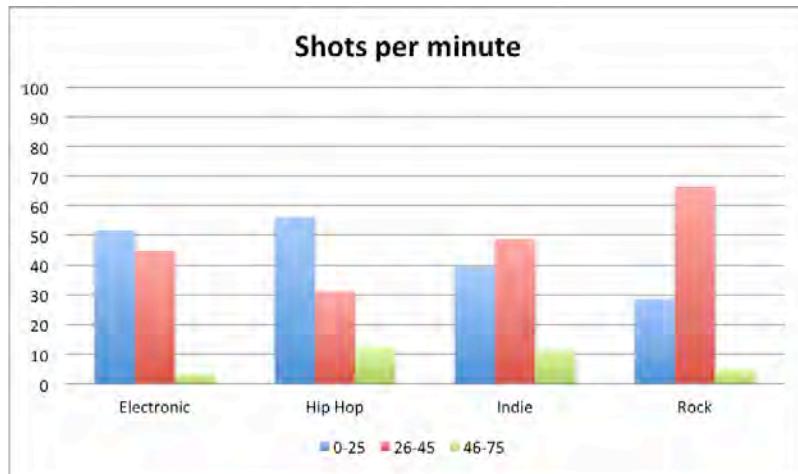


Figure 5.4.9: Shots per minute by genre.

5.4.3 Types of Video Content

The types of video content were openly coded according to the visual material and references frequently occurring in the music videos. The videos were interpreted according to their textual focus, and only those types that were considered essential in describing the video were documented. Four dominant content types, inspired by Lynch (1984), Goodwin (1992), Keazor and Wübbena (2007) and Railton and Watson (2011), emerged from this process: Story topic, dance, performance and abstract/concept. The clips could be described by referring to one or more of these types.

The story topic type indicates that content relates to a story narrative chosen for the video, which may be derived from the song lyrics or a contextualising narrative. The performance type defines a video where the artist performing the song is a central motif in the video. The dance type refers to the highlighting of protagonists or crowds dancing as a central motif. The abstract/concept type refers to videos that foreground a conceptual aesthetic, often in a mostly non-linear or non-narrative fashion that cannot be fitted into any of the other types.

5.4.3.1 Results

While content types still provide for a limited amount of differentiation for categorisation, several strong trends and logical relationships between content types and editing characteristics emerged, indicating that a visual distinction may be more beneficial than an orientation towards music genre.

Values in rhythmic editing, for instance, display an increasing differentiation among content types. Here, the dance type shows a strong orientation towards rhythmic editing, whereas these values are almost completely inverted for the performance type. Figure 5.4.10 also shows that videos classified as containing story topics rarely display rhythmic editing and that concept/abstract videos feature all types of rhythmic editing to a similar extent.

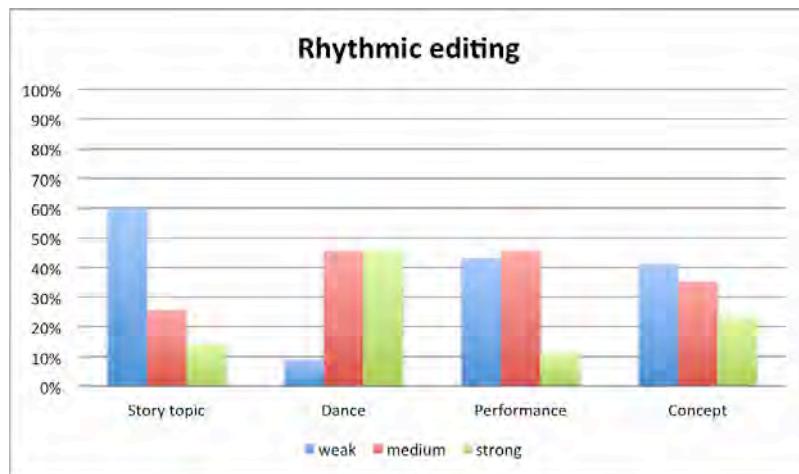


Figure 5.4.10: Rhythmic editing by content types.

A clear distinction also emerged when examining the amount of lip-synching (Figure 5.4.11). While a majority of the videos will feature some form of lip-synching, the differentiation by content type shows more clearly which videos do not feature any visual representation of vocal performance at all. The story topic type, which is based on narrative structure, therefore shows a stronger correlation to videos without this type of performance, whereas videos classified as performance, unsurprisingly always feature lip-synching.

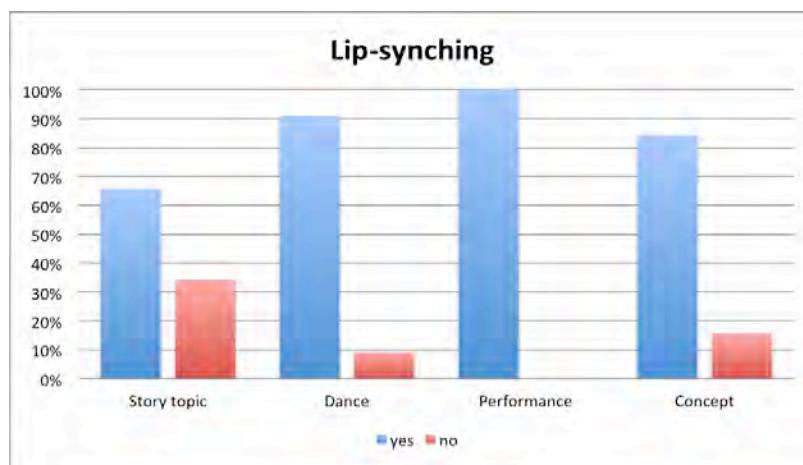


Figure 5.4.11: Lip-synching by content type.

The measure for shots per minute and segmentation, by contrast, deliver similar values and relationships compared to the investigation based on music genres (cp. Figure 5.4.12 and Figure 5.4.13). This may either be attributed to the inability of the classification by content type to provide enough information in conjunction with these characteristics or because these characteristics are not suitable as classifiers in this type of analysis.

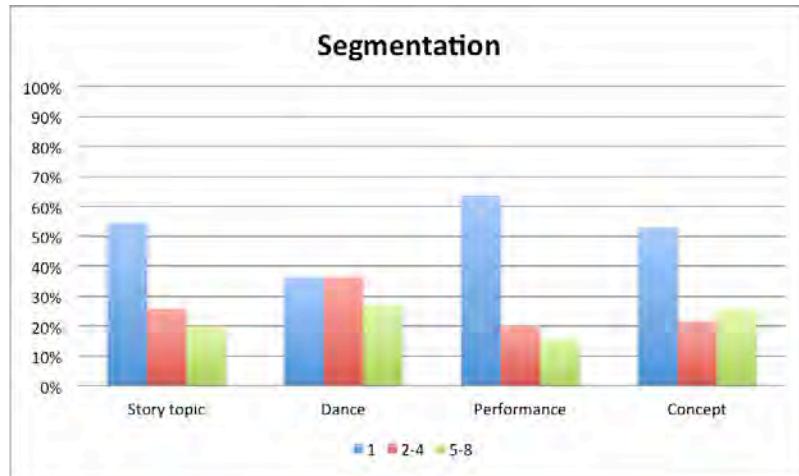


Figure 5.4.12: Segmentation by content type.

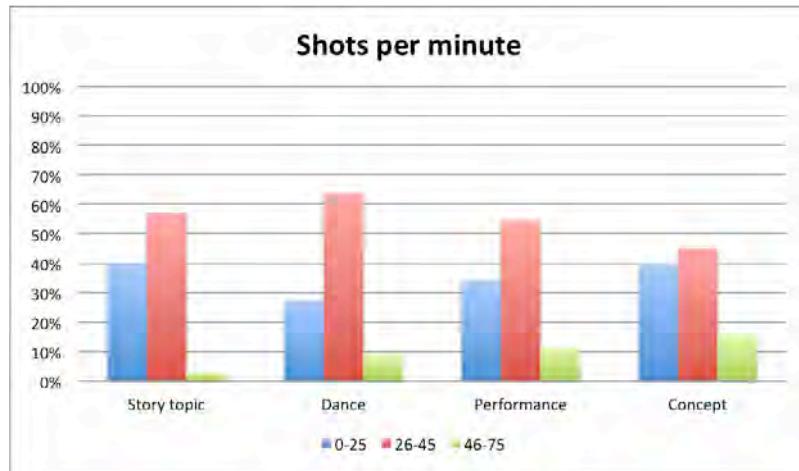


Figure 5.4.13: Shots per minute by content type.

Story topic

Videos classified as story topic have a semi-linear to linear structure with a weak to medium adherence to the song's form, and include little rhythmic editing. The story topic type displays no strong trend in segmentation and lip-synching, and features only a small amount of visual references to lyrical content. The average editing pace is slow to medium.

Unsurprisingly, the story topic content type exhibits a higher percentage of linearity and orientation to the song's form (Figure 5.4.14). Furthermore, it shows the smallest amount of variability in video composition.

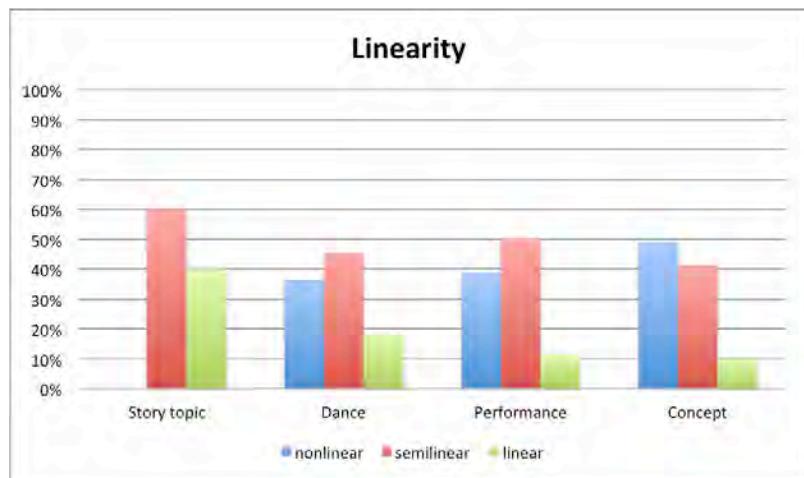


Figure 5.4.14: Linearity by content type.

A video classified as only featuring the story topic content type was *Big Jet Plane* by Angus & Julia Stone.¹¹³ The video follows a simple semi-linear narrative, in which a young woman is shown working as a shop assistant (cp. Chapter 5.2.3).

Dance

Most of the dance videos are non-linear to semi-linear and have a strong tendency towards rhythmic editing. These videos strongly adhere to the songs' sectional divisions. They display the lowest percentage in weak adherence to song structure, as well as the highest percentage in both medium and strong adherence (cp. Figure 5.4.15). The dance type features a larger number of segments than other content types. A majority of videos with this content type feature lip-synching with an overall small to medium amount of visual references to lyrical content. The latter is more dominant in the dance type than in all other types examined in this analysis (cp. Figure 5.4.16). The average editing pace tends to be either extreme, slow- or fast-paced.

¹¹³ NettwerkMusic. “Angus and Julia Stone - Big Jet Plane [Official Music Video]”, YouTube video, 4:05, 30 March 2010, <https://www.youtube.com/watch?v=yFTvbcNhEgc>.

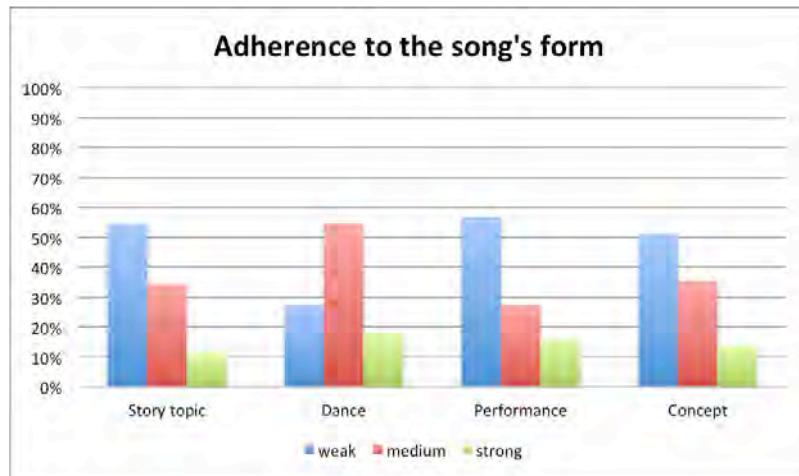


Figure 5.4.15: Adherence to the song's form by content type.

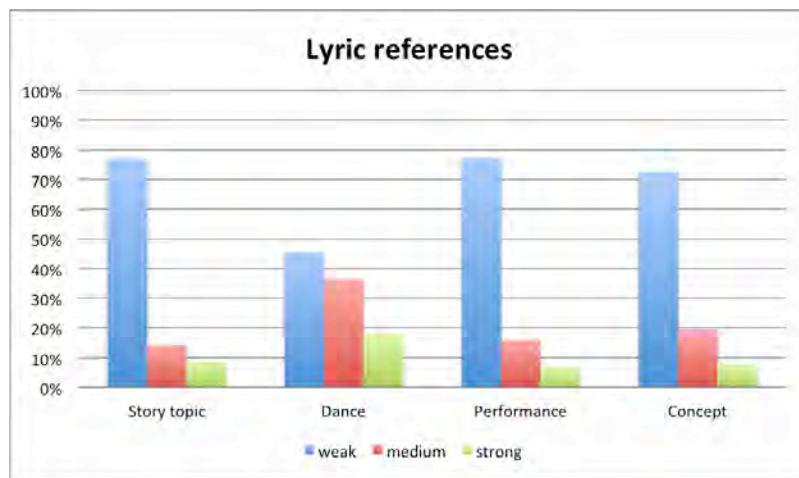


Figure 5.4.16: Lyric references by content type.

The dance type offers a lot of variability with regard to narrative structure but requires a stronger orientation towards musical rhythm. It usually appears in combination with any other type. One music video classified as dance was *Dance the Way I Feel* by Où Est le Swimming Pool.¹¹⁴ This video was also classified as performance. The visual content in this video focuses on the creation of rhythm and a conceptual look and feel for the band. There is no logical progression that would indicate a narrative. Instead, the video develops by increasing the amount of effects, as well as building tension around certain sections of the song.

¹¹⁴ Scorpio Music. “Où Est le Swimming Pool “Dance The Way I Feel” (Alternate Air Version) Official Video”, YouTube video, 3:20, 9 July 2010, <https://www.youtube.com/watch?v=NzWNlZxirQo>.

Performance

Most of the performance videos are non-linear to semi-linear with a weak to medium adherence to the song's form and a weak to medium trend towards rhythmic editing. A large number of these videos show either no segmentation or a structure with only 2-4 segments. All of these videos feature lip-synching with few visual references to lyrical content. The average editing pace is low to medium.

The performance type features a lot of synchronisation to the music but also offers variability in the use of other imagery in the video.

A video classified as performance was *Bang Bang Bang* by Mark Ronson and the Business Intl.¹¹⁵ In this video, we see the musicians in an abstract performance space. The entire video is centred on this performance and supported by dance scenes and visual effects.

Abstract/Concept

Most abstract/concept videos are non-linear to semi-linear with a weak to medium adherence to song structure. The abstract type shows no strong trends in rhythmic editing and segmentation. A majority of these videos feature lip-synching with few visual references to lyrical content. The average editing pace is low to medium.

Characteristic tendencies for the abstract type may vary greatly and offer a high degree of variability.

An example for the abstract content type is the music video for *Plans* by Birds of Tokyo.¹¹⁶ In this video, the viewer may get the impression of a narrative, but the linking of events depends more on personal interpretation rather than a clearly observable logic. The video focuses on the creation of a particular aesthetic in which the band appears as protagonists, rather than performers.

¹¹⁵ MarkRonsonVEVO. "Mark Ronson, The Business Intl. - Bang Bang Bang", *YouTube* video, 5:20, <https://www.youtube.com/watch?v=TM6TCGltfHM>.

¹¹⁶ birdsoftokyo. "Birds of Tokyo - Plans (Official Video)", *YouTube* video, 3:39, 22 June 010, <https://www.youtube.com/watch?v=jSUPaTF-wN8>.

5.4.4 Levels of Narrative in Music Video

At the beginning of this section, the role of narrative and definitions of temporal progression with regard to music video structure was discussed and compared to narrative fiction film throughout subsequent sections. In addition, it was shown that music genre and broad descriptions of music video content provide limited information for the categorisation of music video.

For this reason, an additional classifier for music video analysis and categorisation was devised, largely based on concepts established by Branigan (1992) and Vernallis (2004). This classifier constitutes a spectrum ranging from highly narrative to non-narrative structures. By examining the narrative qualities of each music video and placing them on the narrative scale, it was possible to deduct areas of interest where a clustering of patterns was observed.

(1) Clear narrative

A video with a clear story as central theme. This story displays a structure with a beginning, middle and end. A simplified structure of Todorov's stages of narrative can be applied: equilibrium, disruption, progression and resolution.

(2) Disrupted narrative

A video with a more or less clear story as central theme. Here, we find videos with a narrative structure that can still be perceived as a story (characters act according to certain motivations, visible progression of actions, etc.) but without a clear beginning, middle and end and only parts of Todorov's stages of narrative.

(3) Conceptual narrative

These videos display traces of a narrative, but do not tell a story. Instead, they may be described as the manifestation of a thought, feeling and/or idea. For example, a girl is trapped under frozen water, a character's birth is symbolically visualised, or a fight is staged in an abstract setting. What these videos have in common is that they appear to tell a story through associative connections evoked in the viewer due to the music or expressive visual aesthetics. Typically, the musician(s) will be shown performing in the same setting as the action, serving as an extension to this vague narrative.

(4) Catalogue

In this category, a general topic serves as the central theme around which footage is collected and edited in a non-linear fashion. Often these videos will appear not to be planned or staged. Usually they are neither narrative nor highly conceptualised.

(5) Performance show

These videos provide documentation of a musical performance in a natural space, e.g., a concert hall or rehearsal space.

(6) Visual-musical

Videos with a highly aestheticised conceptual theme but without a narrative progression.

These videos foreground both the song and the musicians who may, for instance, be seen performing in an abstract setting, supported by visual effects.

(7) Visual-conceptual

These videos typically show the progression of a visual theme. They do not focus on musicians or characters but on the development of a certain action, for example, on the creation of an artwork to which the song acts as a soundtrack.

5.4.4.1 Results

The narrative levels described above already point to certain characteristics of content and progression in the way they were defined. The primary assumption that this classification would reveal a more detail the relationship between narrative level and editing characteristics was confirmed by the analysis.

In Figure 5.4.17, for example, similar patterns can be observed for the clear narrative (level 1), the performance show (level 5) and the visual concept (7) categories in terms of linearity. These three narrative levels can be classified as highly linear because they depict processes and stories that would either significantly change in meaning, or become unintelligible when altered. Videos classified as containing a disrupted narrative (level 2) or a conceptual narrative (level 3) show some potential in variability, while videos in the categories catalogue (level 4) and visual-musical narrative (level 6) are highly variable.

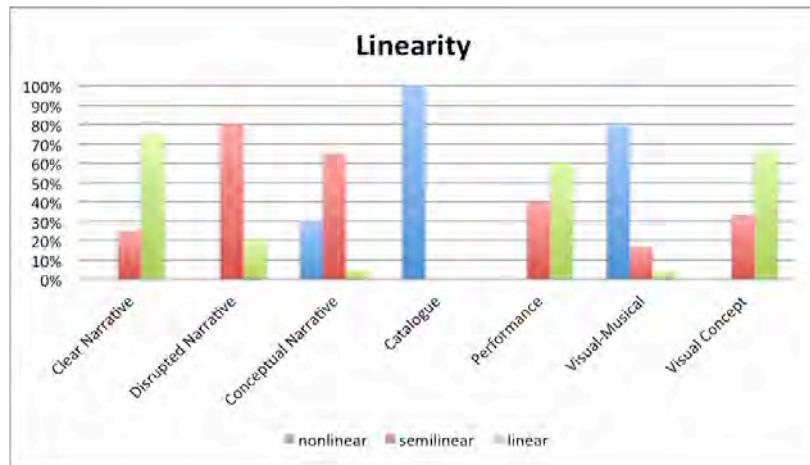


Figure 5.4.17: Linearity by narrative level.

Another logical distinction of shot and editing characteristics can also be observed in Figure 5.4.18, which shows the percentage of videos that feature lip-synching by narrative level. Here, a more detailed overview can be given. The first three narrative levels show a gradual decrease in the number of videos that do not feature lip-synching, whereas the following three levels (catalogue, performance and visual-musical) always feature lip-synching. Visual concept videos, in turn, never show performance.

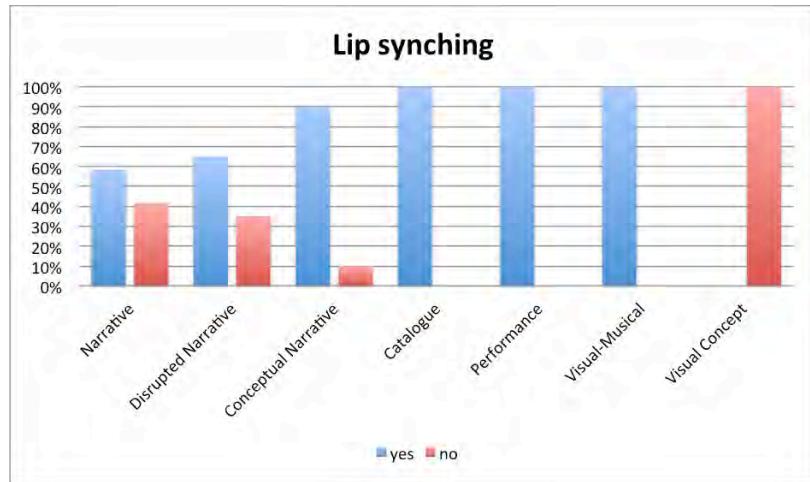


Figure 5.4.18: Lip-synching by narrative level.

Concerning segmentation, again, no strong correlation between this characteristic and the narrative categories could be found. Figure 5.4.19 illustrates that videos in the visual concept category (level 7) never showed segmentation. While the linear progression of this narrative level is a central aspect of its definition, it is possible that a different or larger body of music videos may reveal that other videos falling into this category consist of a small number of

segments and that this category therefore displays a pattern more similar to the other narrative levels in this analysis.

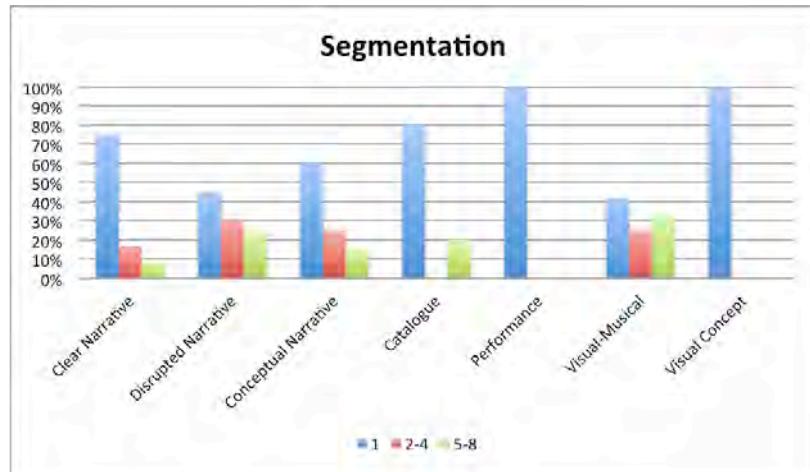


Figure 5.4.19: Segmentation by narrative level.

Visual concept videos in this body of music videos were either realised within one camera take or a small number of shots presenting a continuous process. Therefore, this narrative level also shows a distinct difference to other levels concerning the average editing pace. In Figure 5.4.20, we can also observe that catalogue (level 4) videos were generally more fast-paced than other videos.

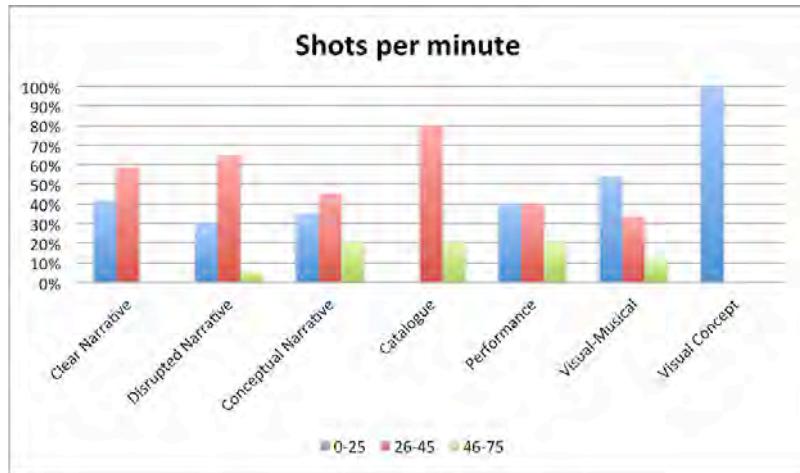


Figure 5.4.20: Shots per minute by narrative level.

Clear narrative (Level 1)

Most of these videos are linear with a weak to medium adherence to the song's form and an either weak or strong rhythmic editing. A large number of videos show no obvious

segmentation. The clear narrative displays no strong trend in lip-synching, as well as a low amount of visual references to lyrical content. The average editing pace is slow to medium. All videos on this level are equivalent to the narrative content type.

The clear narrative builds on a low variability in its narrative structure but a high degree of variability in its adherence to various aspects of the song.

An example for the clear narrative level is the music video for *Teenage Crime* by Adrian Lux.¹¹⁷ This song purely focuses on a single storyline, in which a woman can be seen escaping her day-to-day life. The video starts by showing a presumably typical night in her home, where she tucks in her children before she and her husband go to bed. As soon as he is asleep, she sneaks out to go to a nightclub where she meets a group of young men. Eventually the video suggests the passing of the night, while she has sexual relations with one of these men. In a short sequence the video cuts back and forth between her actions and her family waking up. The story ends with the woman returning home.

This video was classified as containing a clear narrative because it shows many of the necessary characteristics in a short-cut form. The woman is shown to be a traditional mother and housewife. Already within the first seconds the viewer can grasp the impression of boredom and longing, which will soon lead to a disruption. The audience is made to assume that she is not happy and wants to escape, even if just for one night. Eventually she fulfils her wish and enjoys herself with alcohol, dancing and sex. The resolution is kept open. She simply returns home and sits down in front of her house.

Disrupted narrative (Level 2)

All of the videos in this category are semi-linear to linear. The data reveals no strong trends in adherence to the song's form, segmentation or lip-synching. There is a weak to medium amount of rhythmic editing and few visual references to lyrical content. The average editing pace is slow to medium.

¹¹⁷ Ultra Music. "Adrian Lux - Teenage Crime", YouTube video, 2:49, 17 February 2011, <https://www.youtube.com/watch?v=86vDxRr2ELA>.

The disrupted narrative offers some degree of variability in its narrative structure. It is more likely to feature a visual performance and may, therefore, require significant synchronisation of song and visuals.

A video classified as containing a disrupted narrative is *Big Jet Plane* by Angus & Julia Stone¹¹⁸ (cp. Chapter 4.2.3 and Chapter 4.4.3). Although the video features an observable narrative, the viewer receives less information than in the clear narrative.

Conceptual narrative (Level 3)

Most of these videos are non-linear to semi-linear with a weak to medium adherence to the song's form and little to medium rhythmic editing. A large number of videos show none or few segments. A majority of these videos feature lip-synching with a low to medium amount of visual references to lyrical content. The conceptual narrative category does not show a strong trend in average editing pace.

The conceptual narrative has a minimal degree of narrative progression and therefore offers a high degree of variability for video composition. Synchronisation to the song, with regard to the song's form, rhythm and lip-synching seem to be important in this narrative level.

An example of a music video classified as containing a conceptual narrative is *Rock It* by Little Red.¹¹⁹ This video focuses on musical performance and people dancing to the music. As opposed to *Dance the Way I Feel*, for example, the dancing in this video is less stylised and used to symbolise the narrative idea of dancing or celebrating. The video starts out in an abstract Fantasy-like space probably at night time and then moves into a more realistic outdoors area in the morning. The video's concept is based on the feeling of being young and having fun all night long. The overall narrative progression is abstracted in order to translate this concept.

¹¹⁸ NettwerkMusic. "Angus and Julia Stone - Big Jet Plane [Official Music Video]", YouTube video, 4:05, 30 March 2010, <https://www.youtube.com/watch?v=yFTvbcNhEgc>.

¹¹⁹ LiberationMusicAus. "Little Red - Rock It", YouTube video, 3:28, 17 August 2010, <https://www.youtube.com/watch?v=VoQwwcf2ERU>.

Catalogue (Level 4)

All of the catalogue level videos are non-linear and show a weak to medium adherence to the song's form as well as a weak to medium degree of rhythmic editing. A large number of videos show either none or a higher number of segments. All of these videos feature lip-synching and have either a very high or very low frequency of references to lyrical content. The average editing pace is medium to fast.

Catalogue videos offer a high degree of variability due to their lack of narrative progression. The editing pace is faster compared to other narrative levels.

An example for the catalogue level is *Barbra Streisand* by Duck Sauce.¹²⁰ This video consists of a collection of various video materials without any particular narrative or obvious logical progression. The footage was apparently collected according to particular keywords or ideas, such as New York, DJs and fun, and that eventually the video was cut to show the most humorous and visually pleasing shots in accordance to the musical rhythm.

Performance Show (Level 5)

These videos are semi-linear to linear with weak to medium levels of rhythmic editing. Performance show videos stand out in that they do not adhere to the song's form and exhibit distinct visual changes during the course of the video (cp. Figure 5.4.21). None of these videos are segmented and all of them feature lip-synching, but there are no references to lyrical content. Performance videos also show no strong trends in average editing pace.

120 Spinnin' Records. "Duck Sauce - Barbra Streisand (Official Music Video)", YouTube video, 3:16, 5 October 2010, https://www.youtube.com/watch?v=uu_zwdmz0hE.

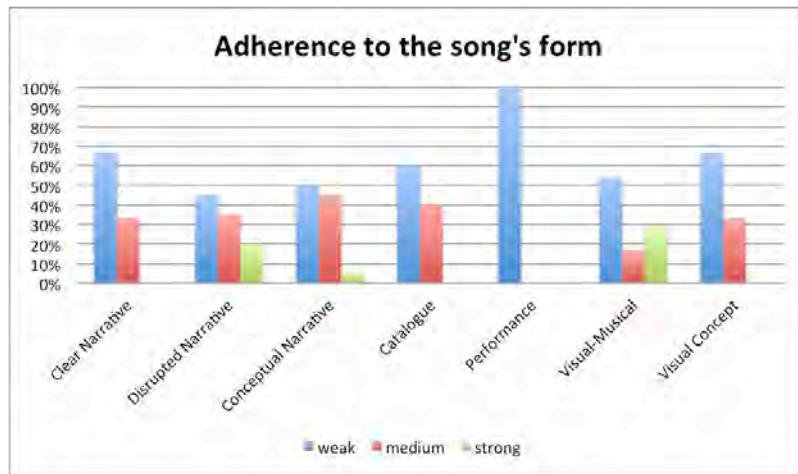


Figure 5.4.21: Adherence to the song's form by narrative level.

Being mostly a documentation of live indie and rock performances, such videos offer little variability. Synchronisation, unsurprisingly, is a dominant feature.

An example for a performance show video with a high level of adherence to the song's form is the music video for *You Got the Dirtee Love* by Florence + The Machine feat. Dizzee Rascal.¹²¹ The video shows the live performance of this song.

Visual-Musical (Level 6)

Most of these videos are non-linear and may have either a weak or strong adherence to the song's form. Visual-musical videos are inconsistent in their rhythmic editing and segmentation. All of these videos feature lip-synching with a weak to medium amount of visual references to lyrical content. The average editing pace is slow to medium.

This narrative level features a high degree of variability. Synchronisation is an important characteristic for lip-synching and orientation to the song's form and rhythm.

A video classified as visual-musical is *Dance the Way I Feel* by Où est le Swimming Pool¹²² (cp. Chapter 4.4.3).

121 BRIT Awards VEVO. “You Got The Dirtee Love (Live at The BRIT Awards 2010)”, YouTube video, 3:38, 29 February 2012, <https://www.youtube.com/watch?v=Zixfls232M>.

122 Scorpio Music. “Ou Est le Swimming Pool “Dance The Way I Feel” (Alternate Air Version) Official Video”, YouTube video, 3:20, 9 July 2010, <https://www.youtube.com/watch?v=NzWNlZxirQo>.

Visual Conceptual (Level 7)

These videos are semi-linear to linear with a weak to medium adherence to the song's form, but a distinctly strong trend towards rhythmic editing (cp. Figure 5.4.22). None of these videos are segmented and none of them feature lip-synching. Considering that the visual concept category featured the smallest amount of videos, it is important to be careful when analysing these results. Some trends may be significantly influenced by the characteristics of individual videos. Although the visual concept videos showed a larger percentage of visual references to lyrics (cp. Figure 5.4.23), it is unclear if this is a general characteristic. The average editing pace of these videos is slow.

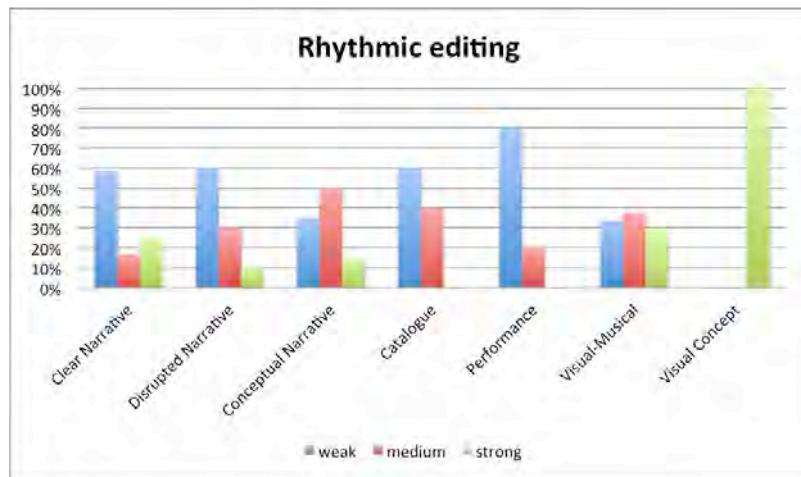


Figure 5.4.22: Rhythmic editing by narrative level.

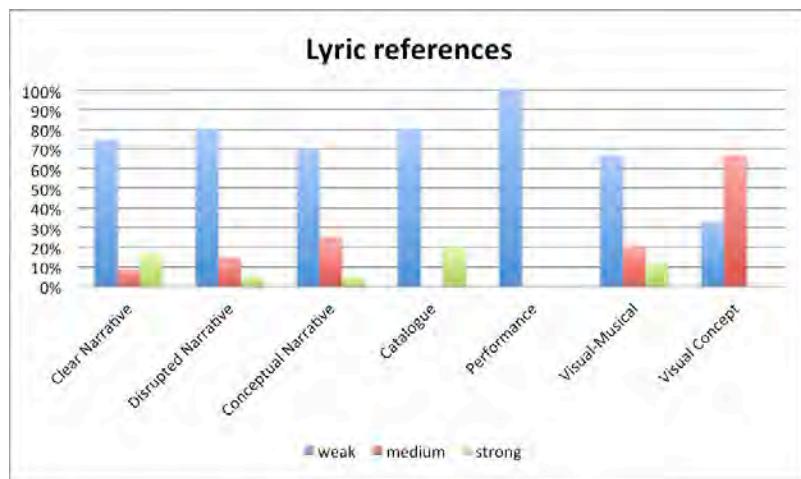


Figure 5.4.23: Lyric references by narrative level.

Visual conceptual videos offer very little variability in the editing process because they tend to display one continuous action. They are strongly orientated to the song's rhythm but don't require lip-synching.

An example for a visual concept video is the music video for *Addicted* by Bliss N Eso,¹²³ where the viewer can see the progression and movement of graffiti artwork in a stop-motion animation.

5.4.5 Discussion

The videos analysis presented in this section provided a broad view on music video aesthetics and the context in which they operate. This analysis strongly focused on technical characteristics, which may become relevant for the development of generative editing algorithms and future implementations of templates and frameworks. In this context, the first attempts in finding methods of categorisation and defining music video aesthetics focused on broad classifications that did not consider dynamic progression within a sequence.

Because this approach is based on technical editing characteristics rather than a more comprehensive view of visual content, some of the results may be surprising, e.g., the general lack of strong coherence with the musical form and elements of a song. The analysis did not consider the human know-how and intuition required to influence the perception of what looks and feels right to a viewer. For instance, an editor might refine a cut to one or two frames in order to create a particular effect. This type of precision is necessary to control the perception of synchronisation, continuity, and discontinuity in a viewer by balancing audio-visual rhythm according to both content and cut-point.

Three different classifiers for categorisation were tested and reviewed in this analysis to evaluate their practicality and applicability. The classifiers consist of compound characteristics in varying levels of complexity. These classifiers are: music genre, video content types, and a seven-level narrative spectrum.

Video content types served as the first visual classifier in this analysis and are based on a comparatively simple descriptive model containing four combinable visual characteristics, i.e.,

123 LiberationMusicAus. "Bliss N Eso - Addicted", *YouTube* video, 3:10, 7 November 2010, <https://www.youtube.com/watch?v=tDaFTVb662E>.

dance, performance, story topic and visual concept. Based on previous music video research, these characteristics denoted common themes in visual content. Accordingly, the visual content type analysis provided some information on sequence composition and can therefore serve as an analytical tool that may function as a general guide in music video classification.

For a classification with a particular focus on editing techniques, the narrative progression was foregrounded and a seven-level scale was defined. The seven levels on this scale represent the amount of rigidity and flexibility of a music video's narrative structure. Technical parameters outlined within the analysis correlate more closely to these levels and therefore provide a possible framework for classification.

The results demonstrated that a comprehensive analysis can help to establish systems of reference for music video aesthetics. However, they also suggest that it may be neither useful, nor possible to develop definitive templates for the editing of music videos beyond comparatively broad descriptions or parameter models that may serve as starting points for further editing because of the wide variety of music videos that seek many different kinds of audio-visual relations and the resulting flexibility that directors require to create these videos. A software tool for generative editing therefore needs to be orientated towards providing editors with a maximum amount of control and creative freedom during the editing process. Music video is known for its evolving aesthetic and the pushing of boundaries in audio-visual composition for entertainment and advertising. Thus, few music videos fit into strictly defined categories and structural definitions, since content and progression depend on various factors, which, from a technical point of view, may not be classifiable in their entirety. The development of a software tool for generative editing should therefore be based on methods that maintain the potential for innovation and experimentation.

While the analysis has confirmed the difficulty of clearly categorising music video, as outlined at the beginning of Chapter 5.4, the results suggest that a suitable categorisation can be achieved when focusing on specific goals. The results further suggest that it is preferable to use video-specific multi-dimensional descriptors over single parameters for music video classification in order to arrive at a certain level of relevance and accuracy across a large body of music videos.

The analysis further confirmed assertions by Jones (1988) and Railton and Watson (2011) that music video genres have to be defined within their own right and cannot simply be obtained

by adopting classifications based on music genres. In this way, music video appears to establish its own form parallel to the music that it supports.

For a future implementation in a potential software tool, the results can serve as a useful starting point, for instance, by creating provisional templates, which allow editors to quickly view differently edited sequences in a rough-cut form, in order to explore some of the music video's potential, as well as possibilities that might not have been considered yet. With this approach, the editor will be free to keep, transform and add onto any suggested values.

The following chapters will discuss the practical and experimental work undertaken as part of this research project. The creation of generative editing software was influenced by this analysis in so far as the analysis dictated several conceptual requirements for further development. The algorithms were designed to be highly modular and variable with regard to music video's characteristics identified as important for its progression. Therefore, the following experiments were also used to explore different methods of temporal composition that influence both the creation and final aesthetics of the resulting generative music video sequences.

6 Music Video Editing and Structure

“To understand editing, we must understand it as a form of communication, without attempting to reduce it to a model of writing and reading.”

(Carroll 1996, 403)

Walter Murch suggested that cinema may be in a stage similar to music before notation, i.e., a process that allowed for the deeper analysis and formalisation of musical practice (Ondaatje 2002). At this point in time, it is hard to predict if a cinematographic or videographic notation of any form will be feasible. However, with the necessity of formalising some processes of traditional editing for generative video editing practice, this possibility should be considered.

The development of *Dividation* itself constitutes a method of learning how to formulate editing processes, in addition to learning the potential practice of generative editing. This process was supported by the investigations into music video aesthetics in Chapter 4, as well as the analyses in Chapter 5, in order to derive conceptual definitions of music video, editing dynamics, patterns, and functions.

The discussion on the history of music video and its aesthetic qualities established that montage theory might offer some insight concerning music video editing, especially due to their common lack of diegetic sound and the resulting necessity of producing meaning dominantly through images and their juxtaposition in relation to music. Montage theory and the role of the individual shot will therefore be discussed next, before relating this information to the system’s technical requirements in the following chapter.

This chapter will therefore focus on selected concepts deriving from film theory in order to establish an understanding of video structure, and how this structure can be achieved using generative editing methods. The concepts and theories discussed here were chosen according to their relevance for music video, and serve to complete the theoretical repertoire required for the successful application of generative editing, before addressing the development of *Dividation*.

In the previous chapters, film and video editing were discussed with a particular focus on music video. Since editing constitutes a highly intuitive artistic practice, it was established that

the practice of editing is difficult to quantify. Nevertheless, some conventions and recommendations for generalised editing mechanisms could be identified, which will be discussed in this chapter, as well as the perception of edited sequences.

In order to find structural definitions of the cinematic form, film has been analysed in the context of language in the early to mid-twentieth century. Metz (1974, 44) pointed out that this approach has always moved between two positions: “the cinema as a language” and “the cinema as infinitely different from verbal language”.

Eisenstein spoke of editing as a poetic language, which borrows techniques from literature (Bordwell 2005). Metz (1974) also approached cinema in terms of language, interpreting the Saussurian concept of the sign to be translinguistic. Metz highlighted the absence of words in this cinematic language, instead referring to the individual shots as its sentences. He claimed that the cinematic image is of an assertive quality, or, in other words, it is always actualised, i.e., it will always convey more information than a single word. In this sense, due to its temporal nature, the shot must always depict an event.

Metz (1974, 69) further stated that “cinema can speak only in neologisms” and that every “image is a hapax”. In contrast, Nichols (1975, 33) then claimed that cinema is neither speech, nor language. Rather, according to Nichols, it is “stylistic before it is grammatical” and, considering that film can only be spoken in ideolects, it is comprised of conventions rather than the characteristics of language. Monaco (1981, 172), in turn, spoke of “vaguely defined rules of usage in cinematic language”, emphasising the natural evolution of this kind of syntax as a result of usage and not a determinant.

In summary, the discussion on film grammar suggests that editing cannot be described in the same terms as language. It is an individualistic, evolving practice which Murch referred to as a continuous repetition of decisions regarding what kind of shot needs to be shown next in the sequence, where the shot needs to start, and where it has to end (Ondaatje 2002). Rather than investigating the existence of a syntax and grammar, it therefore may be better to ask what motivates these decisions in the editing process, or, in other words, what makes a cut good, or right.

Several theories and opinions exist on what constitutes an ideal cut. In general, there is no definite set of rules. While Murch mentioned several mathematical editing principles he

discovered over the years, such as the number of thematic elements that can be processed by the audience or the average number of camera positions in different types of scenes, more often editors and authors will refer to intuition and emotion as the most important guidelines in editing (Oldham 1992, Ondaatje 2002).

Film and video editors know, almost intuitively, that some cuts work and some cuts don't. We perceive some scenes as fluid and engaging, others as disruptive and disorientating. Many sequences will rely on fluid and continuous action; others may seek to engage the viewer in different ways.

It is unlikely that we are able to identify a definite universal set of rules on how to compose video sequences that both artists and scholars will agree upon. One presupposition advocated by the majority of film and video editors seems to be that editing, first and foremost, is a matter of instinct and intuition. According to them, it may or may not be good to follow formal conventions, but eventually the process of editing depends on the artist him- or herself, and the material he or she is working on (Oldham 1992, Oldham 2012). We can assume that both a thorough knowledge of editing practices, techniques, film and video formats as well as genres, and experience are essential to this practice. However, according to editors such as Mark Livolsi and Alan Berliner, very often an effective cut also depends on trial and error. These editors also emphasised the importance in finding the "emotional truth" in an edited sequence and engaging the viewer in this emotion (Oldham 2012, 33).

According to Balázs (1970, 131), the "art of cutting consists in the first place in determining the length of each shot". Therefore, "if a shot is a little too long or too short, it can decisively change the effect of the picture, just as a tune is completely changed if a tone is pushed up or down half an interval". He preferred the term "montage" to "editing", because he considered the notion of assembly with regard to cutting a sequence more adequate. "Shots", he elaborated (Balázs 1970, 118), "are assembled by the editor in a pre-determined order, in such a way as to produce by the very sequence of frames a certain intended effect, much as the fitter assembled the parts of a machine so as to turn these disjointed parts into a power-producing, work-performing machine".

With this notion of a pre-determined sequence we may ask what influences the duration of a shot. Most editors interviewed by Gabriella Oldham (1992), for instance, stressed the importance of intuitive knowledge concerning what needs to be communicated to a viewer.

Following this idea, a shot should end when it has revealed everything it could or needed to show, something that editors often describe as the intuitive knowledge when it is over. In addition, Bordwell and Thompson (2010) explained that often the duration of a shot will also depend on its technical parameters in conjunction with the amount of information communicated in it. For example, long shots will often be on screen for a longer time than close-ups.

According to Murch (2001, 18) the ideal cut is based on six criteria:

- (1) Emotion,
- (2) Story,
- (3) Rhythm,
- (4) Eye-trace,
- (5) Two-dimensional plane of screen,
- (6) Three-dimensional space of action.

These criteria are deliberately put in descending order. If necessary, the bottom criteria may be sacrificed in order to save those on top. Murch further suggested that editing should always aim to create the right emotion in order to immerse a viewer, whereas formal aspects such as spatial continuity may be violated when it serves the emotional content of the sequence.

Dmytryk (1984) probably proposed the clearest concept on good editing practice, claiming that the following seven rules must be followed:

Rule 1: *Never* make a cut without a positive reason.

Rule 2: When undecided about the exact frame to cut on, cut *long* rather than short.

Rule 3: Whenever possible cut “in movement”.

Rule 4: The “fresh” is preferable to the “stale”.

Rule 5: All scenes should begin and end with continuing action.

Rule 6: Cut for proper *values* rather than for proper “matches”.

Rule 7: Substance first – then form.

Dmytryk also stressed the necessity of precision. However, similarly to Murch he argued in favour of meaning and “humanity”, foregrounding that editing is not a mere mechanical task,

but a practice that seeks to establish communication with the viewer (Dmytryk 1984, 145). Perhaps it is this notion of communication which exacerbates the difficulty of formalising the practice of editing. We must know in advance what exactly a given sequence of moving images seeks to communicate to its audience when making editing decisions. However, a music video, as discussed in earlier chapters, will usually operate on different motives than a narrative feature film. It is evident that a formalisation of music video aesthetics cannot rely on conventions based on logical continuity, as this device is rarely used in music video.

Editing methods for music video have to strive for a synergy between musical and visual content, but it may not be possible to entirely formalise these audio-visual connections for implementation within an automated software tool either. Furthermore, although highly experienced and skilled editors will be more likely to create works that appeal to a larger audience, the perception of these connections also tends to be highly subjective.

Underscoring this difference, Chion (1994) and Vernallis (2004) compared the nature of music video to early Soviet film and montage theory, pioneered by filmmakers such as Sergei Eisenstein, Lev Kuleshov, Dziga Vertov and Vsevolod Pudovkin. Indeed, several aspects in these cinematographic theories may allow us to distinguish music video editing from film editing, and to assist in deriving some formal descriptions that are helpful for the development of editing algorithms.

6.1 Montage Theory

When speaking of montage in film and video editing, we have to consider the term's wide range of meanings. Commonly the term is now known because of montage sequences employed in many Hollywood films. These sequences are used to stylistically condense time and space through the use of music and discontinuous editing.

Montage originated in the Russian cinema of the 1920s and 1930s. The overarching concept of montage could be described as the “juxtaposition of two or more separate event images that, when shown together, combine into a new and more intense whole” (Zettl 2014, 397).

The Russian formalist filmmakers understood montage as a method of juxtaposing shots in order to assemble a sequence. The focus of montage was less on narrative clarity, but predominantly on concepts such as rhythm, and meaning-making through the use of

metaphoric and mental associations (Bordwell and Thompson 2010). This development was predominantly inspired by visual artists and, subsequently, adopted by filmmakers such as Sergei Eisenstein, Lev Kuleshov, Vsevolod Pudovkin, and Dziga Vertov, who encouraged the use of the term in cinema in the period from the 1920s to the 1930s (Bordwell 1972, Bordwell 2005). These Russian film pioneers were influenced, on the one hand, by both artistic and political developments, e.g., Futurism and Constructivism, and, on the other hand, by the works of the American cinematographer D. W. Griffith, who, following Edwin S. Porter's concept of the shot being the central unit of structure in film, introduced essential methods of dramatic construction to film (Bordwell 1972, Reisz 2010, Dancyger 2011).

Lev Kuleshov referred to montage as the organisation of cinematic material and established awareness of a phenomenon with his experiments, which would later be referred to as the Kuleshov effect (Kuleshov 1974, Cook 2004). Kuleshov's experiments explored and asserted the notion that editing was able to create meaning through the practice itself, rather than just through the content of images. More precisely, the Kuleshov effect suggests that the meaning of a shot depends on its context and that a change in the order of shots may change the meaning of a scene or sequence (Bordwell 2005). Kuleshov further referred to montage as "the essence of cinema technique", as well as "the essence of structuring a motion picture" (Kuleshov 1974, 183).

At the same time, Sergei Eisenstein was particularly active as a film theoretician. A central concept in Eisenstein's theories was the conflict between images. Eisenstein used this metaphor, among others, to discuss both visual rhythm and montage's dialectic quality (Bogue 2003, Cowan 2007). Eisenstein also tried to introduce biological concepts into film theory by referring to shots as "montage cells", arguing for the idea of letting a film grow, much like a living being (Bogue 2003, 50).

The early Soviet filmmakers were therefore among the first to clearly articulate the importance of the single shot and its order in a sequence. Montage was, in this sense, an artistic movement in film making that, although based on narrative structures, tried to transcend narration and naturalism by exploring the possibilities of metaphor and rhetoric (Bordwell 1972).

An important aspect to montage style may have been the fact that synchronised sound had not yet been established in film. It has been repeatedly stated that silent films never really were

silent, but that often sound had been considered only as an accompaniment to the images (Rabiger 1989, Altman 2004, Dancyger 2011). Before cinematic sound recording and design became more sophisticated in the following decades, the images necessarily bore a greater responsibility for expression. Indeed, some filmmakers opposed the introduction of the talkie because it was expected to undermine the purpose of montage and add too much credibility to these images (Dancyger 2011). Montage, therefore, was more concerned with assembling images, and its effect on the relation between image and sound has always been at issue.

6.2 The Shot as Granule

The shot has so far been introduced as the basic element of structure in film. This idea formed the basis of the theoretical and practical context that established film and video editing from the beginning of the twentieth century, when it was introduced by Edwin S. Porter and adopted by other filmmakers, most notably D. W. Griffith. The individual shot has always been of interest anytime film or video structure has been discussed and analysed, first in montage theory, and in the following decades most prominently in theories that foreground linguistics and narratology as models for investigating film and video structure.

For the development of *Dividation*, which is discussed in Chapter 7, the shot is also considered the basic building block or granule from which sequences are created. In order to generate meaningful structures, it is therefore essential to understand how to describe a shot in the algorithmic process. The following theories on the description of shots were selected to provide an overview of potential technical parameters that may be useful in the generative editing process.

In this context, a shot is a single take recorded by a camera, which is cut to depict the relevant events of this recording. The shot can be broken down into frames, i.e., individual images, which, when played in sequence, create the illusion of movement.

Bordwell and Thompson (2010) introduced four different relations between shots created through the practice of editing: graphic, rhythmic, spatial and temporal. It is according to these themes that the editor exercises control over the sequence. Editing then is what “permits the interaction, through similarity and difference, of the purely pictorial qualities of ... two shots” (Bordwell and Thompson 2010, 221).

In order to describe a shot itself, Bordwell and Thompson (2010) considered three essential factors:

- 1) the photographic aspects of the shot,
- 2) the framing of the shot, and
- 3) the duration of the shot.

With regard to technical solutions towards an automated editing, Davenport, Smith and Pincever (1991) suggested to use four descriptors to define a shot:

- 1) Narrative perspective or point of view,
- 2) the relation between camera and sound recorder to scene and environment,
- 3) content, and
- 4) context.

According to Davenport, Smith and Pincever, these descriptors are time-dependent and a complete annotation of footage may require an unreasonably large amount of data in order to accurately describe these shots, depending on both the nature of the footage as well as the requirements for sequence creation.

Snoek and Worring (2005) introduced methods for multimodal indexing of video footage, including edited sequences. By viewing a video as the result of an authoring process, Snoek and Worring proposed three perspectives for indexing videos: semantic, content, and layout. The semantic perspective describes the video's purpose, e.g., entertainment or information, its genre, and depending logical units. The later are introduced as referring to particular types of scenes, or other structural sections that can be identified in a film or video format, for example the sections of sports games or news broadcasts. These logical units constitute named events which identify particular actions, ranging from “explosions in action movies, goals in soccer games, to a visualization of stock quotes in a financial news broadcast” (Snoek and Worring 2005).

For the content perspective, Snoek and Worring emphasised the importance of

- 1) setting,
- 2) objects,
- 3) people, as well as
- 4) visual aspects such as colour, and technical camera parameters, e.g., distance and movement.

Setting closely relates to the above-mentioned logical units. Objects and people in turn are considered the main elements in named events.

The layout perspective refers to the identification of frames, shots and cuts, as well as other work steps carried out in post-production such as the addition of special effects.

Summarising these different perspectives on describing video footage, a detailed shot description should therefore include the following parameters:

- Photographic aspects concerning lighting, colour, and lens optics,
- Camera distance,
- Camera movement,
- Shot duration,
- Narrative or otherwise structural context,
- Location and setting,
- Objects,
- People.

Given this diversity of parametric taxonomies, it seems that the set of parameters suitable for describing a shot also depends on the context in which these descriptions are utilised.

7 Dividation: Generative Music Video Editing Algorithms

“Well, I’ve a hunch that some systematic approach to cinematic notation is possible but has yet to be developed.”

Walter Murch (Ondaatje 2002, 295)

7.1 Background

Before discussing the development of *Dividation*, the technical requirements and concepts regarding its realisation will be reviewed to establish the most important areas of interest for this research phase, and to gain a better understanding of *Dividation*’s scope and basic methods.

7.1.1 Requirements

The development of *Dividation* as a process required careful consideration of both technical and creative concerns. The creation of a video sequence with the help of *Dividation* is based on the idea of screen events unfolding in time. Therefore, an iterative algorithmic decision process is used to assemble a sequence shot by shot. At the beginning, the requirements for algorithmic generative editing were focused on technical specifications in order to establish basic processes for algorithm development. These requirements were defined according to the following five essential features:

- 1) real-time playback,
- 2) synchronisation,
- 3) musical structuring,
- 4) visual- as well as narrative structuring, and
- 5) individual assembly of shots each time the video is viewed.

The requirement of real-time playback of generative music video sequences was an essential aspect in the development of *Dividation* because it allows for the individual assembly of video sequences and a potential adaptation during playback, with the possibility to add interactive features in the future. However, the real-time application of generative editing methods also constituted one of the greatest challenges during the experiments. It imposed a major

constraint with regard to the time allowed to process algorithmic editing decisions, which may have a negative impact on computing performance and download speed.

These challenges are also closely related to the issue of accuracy, required for audio-visual synchronisation, most notably for shots that depict musical performance, particularly lip-synching. These shots can only be played at one specific moment in the video. The design of *Dividation* therefore needs to include an efficient method for managing different types of footage, which can accommodate temporal accuracy when required. This includes a consistent system for the preparation of footage, an adequate file structure, and documentation of metadata including shot durations and start times, as well as identification of the different types of footage and an efficient handling during the algorithmic decision process.

The technical experiments with *Dividation* did not include a detailed musical analyses, e.g., beat and key detection. The music video analysis in Chapter 5.4 has shown that music video editing rarely displays an excessive metrical accuracy. For example, Tom Haneke argued that editing sequences to music could not be approached systematically. Cuts may therefore be placed anywhere around a beat depending on the editor's intuition (Oldham 1992). The reason for this openness with regard to audio-visual rhythm can be attributed to the combination of visual and sonic features, the order of footage in a sequence and the viewer's perception of these combined factors. The extent to which this perception of rhythm can be influenced by using algorithmic methods remains a topic for future investigation.

Audio-visual synchronisation, however, is maintained because it is connected to the video's inherent visual logic or narrative progression, which in many cases may be completely separated from the musical rhythm, leading to different types and functions of footage, as well as different methods for their structuring in a sequence.

The structuring methods available in *Dividation* essentially define the music video's progression and thereby a large part of the editor's creative choices. The development of generative editing methods implies to determine how editors will interact with such an editing system, how the progression of a sequence can be controlled, and how their creative practice will be transformed in comparison to traditional forms of film and video editing. In the case of *Dividation*, generative editing will be more analytic, and therefore potentially more explicit, than traditional editing.

Dividation uses an editing algorithm based on Markov chains, utilising relationships between annotated visual characteristics in the video to progress a sequence. The following section, which discusses the technical realisation of *Dividation*, will further explain the choice of this method.

The editor must be able to analyse and describe the video's logical or narrative structure according to these characteristics. Figure 7.1.1 shows an example of visual characteristics used to determine the editing dynamics of a music video sequence in addition to other characteristics.



Figure 7.1.1: Example of visual characteristics in a music video shot.

Therefore, for the development of *Dividation*, both the technical realisation of the system, guided by viewer experience, and the editor's interaction with the system had to be considered.

For the generation of adequate sequences, a particular challenge in the development of *Dividation* is the maintenance of a fine balance between rigidity and variability. If the rules are too rigid, the individual video sequence iterations will become too similar, defying the purpose of developing algorithms for generative video editing. However, rules that do not provide enough structure for algorithmic decisions are problematic as well because they may result in sequences that are perceived as random and disjunctive, and therefore most likely lead to the disengagement of the viewers.

With *Dividation* two methods of cutting were tested, and therefore different types of footage were required to be supplied to the editing software. For the first method, sections of footage of varying durations were provided and then an algorithm chose a section and therefore dynamically cut the provided footage. The second method required the editors to manually cut a number of shots to provide start and end points, removing the cutting function from the software. In Chapter 7.2.2 and Chapter 7.2.3, these methods are discussed in more detail.

During the development of *Dividation* it became clear that it is essential not to provide all possible information of a shot at once but only those parameters differentiating the shots in a particular context. This constitutes an important aspect of describing shot content and defining sequence structure. It was already suggested that generative editing practice may be a more rational process compared to traditional editing. The editor must be able to both synthesise essential visual features in each shot and to place them according to their logical relationship in order to describe the dynamics of a video sequence, and therefore communicate this vital information to the editing algorithm. The editor may already be aware of these elements and use them purposefully to construct a sequence, or he or she may be using these elements intuitively. In *Dividation*, these elements and their rationalisation constitute an essential means of control over the process.

For this reason, during the development of *Dividation* different methods for structuring a sequence were explored, from editing methods dominantly driven by shot annotation to list-like structures providing rough guidance for the sequencing of events.

Any definitions for these structuring methods must probably be derived individually for each generative music video because different music videos may follow vastly different concepts and therefore require a prioritisation of different visual features and structures. For instance, some sequences may highlight particular sections or moments in the video by changing the camera distance while structuring the overall progression according to locations and colours. Other sequences may emphasise the appearance of protagonists and objects in the setting. Because music videos also feature a wide range of complexity from a narrative and visual point of view, some videos may rely on few visual features, whereas others require a more detailed sequence of events.

With the shot as the basic building block, generative editing becomes a reiterative decision process structuring shots according to visual features or narrative structures putting these shots into a temporal order.

7.1.2 Technical Realisation

Despite the previous attempts to outline music video aesthetics and editing dynamics described in Chapters 4 and 5, as well as the search for formal definitions regarding editing structure described above, we still don't know how film and video editing really "work" and we were not able to derive a formal definition for this practice. It becomes clear that we cannot simply begin by feeding a computational system information on how to edit a music video sequence. Instead we have to learn how to approach this task by combining both theory and practice as a basis for the iterative practice-led enquiry enabled by the development of *Dividation*.

By teaching *Dividation* how to edit, we can learn what editing means to a generative system and what it takes to make a music video sequence look and feel right to the viewer. *Dividation* avoids being overly prescriptive in order to still allow the editor to explore his or her creative possibilities. The software therefore aims to support the creative interaction between algorithm and editor as a kind of creative partnership between human being and technology (Brown 2012), and to assist in gaining a better understanding of our creative abilities as we decide which aspects of our editing practice are best to delegate to the software.

Given the experimental and provisional nature of *Dividation* as a research tool, the outcomes of *Dividation*'s development will qualify as prototypes rather than fully functional software programs.

The previous chapters raised a number of questions important for the development of *Dividation*. First, from an editor's point of view, the following questions need to be answered:

- How can editing aesthetics and conventions as well as any given editor's creative style be synthesised by the algorithm?
- How can users interact with a system for generative music video editing?
- How can the user's creativity be maintained or enhanced and what does it mean to be creative when using *Dividation*?

And second, from a technical point of view, the following issues arose:

- How can issues of control and variability be balanced by the system?
- What makes a sequence created with generative editing “right” or “wrong”?
- What is required for the system to function in real-time?
- How and to what extend are author and audience involved in the creation process?
- How explicit are *Dividation*'s definitions for generative editing?

These questions will be addressed alongside core issues concerning the development of generative algorithms, the development of user interfaces, and the technical realisation of a software implementation on an online platform.

Before describing the design of these software tools, it was necessary to formulate the algorithmic process itself and thereby establish how the system will create video sequences in the first place. This required defining both the way in which these sequences could be assembled by the computational system and an adequate workflow translating the author's creative concepts for this purpose.

7.1.2.1 Implementation

Dividation utilises a stochastic process based on Markov chains which is defined by the definition and succession of a specified number of states in order to create sequences. A transition probability is allocated to each step in the succession of these states. In so-called first order Markov chains, which will be of primary interest in this thesis, these transitions only depend on the immediately preceding step (Grinstead and Snell 1997).

The use of stochastic algorithms and Markov chains is a common method for generative music composition (McAlpine, Miranda and Hoggar 1999). Video sequences can be considered to consist of sequences of shots, with each sequence requiring an individual editing decision. Therefore, for the development of *Dividation*, some of the concepts used in generative music composition were transferred to generative video editing.

Considering its purpose as a creative tool for the music video editor, software intelligence in *Dividation* has to remain limited (cp. Chapter 3) compared to other programs, for instance, AARON (McCorduck 1991). *Dividation* is not meant to exercise independent creative acts but to provide the function of variability to the editor who wants to create variable videos.

Therefore, while the machine will be responsible for finally assembling the sequences, the authority over progression, look and feel remains with the director and has to be individual for each generative music video created with the software. A general development of definitive templates that can be applied to a body of video footage was therefore not considered as appropriate. Instead, an interface to the algorithm was developed exposing salient parameters to the author's control. Stochastic elements in the algorithm allow for variations in each generated video within the boundaries set by the user.

A central consideration for generative music video editing with *Dividation* is its application in online media, as described in Chapter 2. The *Dividation* system has therefore been developed in HTML5 using a MySQL database. PHP scripts were used to assemble the video sequences, and playback functionality was supplied with JavaScript in conjunction with the HTML5 features 'video' and 'canvas'.

Other software environments tested during development were openFrameworks using C++ and Microsoft WebMatrix using Razor and C#. The combination of HTML5 with JavaScript, MySQL and PHP was chosen over other approaches in order to make use of open-source technologies in a simple setup that supports direct implementation on a website as well as cross-platform compatibility during development.

7.1.2.2 Algorithm Development

Before discussing the specific software experiments, this section will briefly describe several basic concepts relevant to the processes developed in *Dividation*.

These concepts concern the editor's control over sequence outcomes through the timing of algorithmic decisions, cutting methods, and different approaches regarding structural definitions. By using computational automation and randomness, questions of control must be highlighted and prioritised in this research. During the development of *Dividation*, the role of the editor was always the centre of attention because, according to Candy and Edmonds

(2002, 27), “the artist must be able to decide on the issue rather than have it decided by the technology”. Any methods developed in this process therefore had to be able to accommodate his or her practice and ensure possibilities for creative input.

Two basic approaches were adopted to define the timing of the decision process. They can be differentiated according to the amount of control they enable over the sequence outcomes and their performance.

The first method generated sequences during playback, which means that the decision for placing of a certain shot in the sequence was made shortly before it was shown to the viewer (cp. Figure 7.1.2). While this process provides the highest potential for variability, it also inevitably leads to a decision process that cannot be corrected afterwards, which makes the planning and creation of a coherent video structure comparatively risky or difficult compared to non-real-time editing. Additionally, the support of real-time editing also exacerbates technical issues such as file loading times and uninterrupted playback.

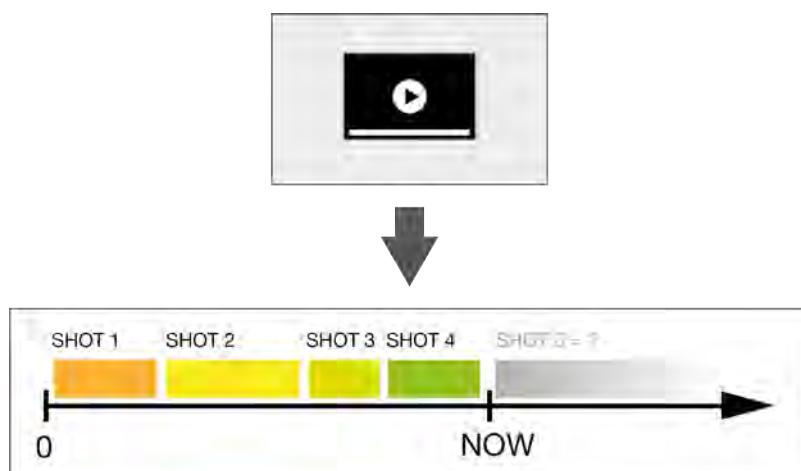


Figure 7.1.2: Illustration of the real-time assembly of music video sequences.

A second method was then designed to address some of these issues by optimising buffering times and therefore potential waiting times for the viewer. Instead of assembling the video sequence during playback, the sequence is planned and an edit decision list (EDL) is generated before the generative video is played back to the viewer (cp. Figure 7.1.3).

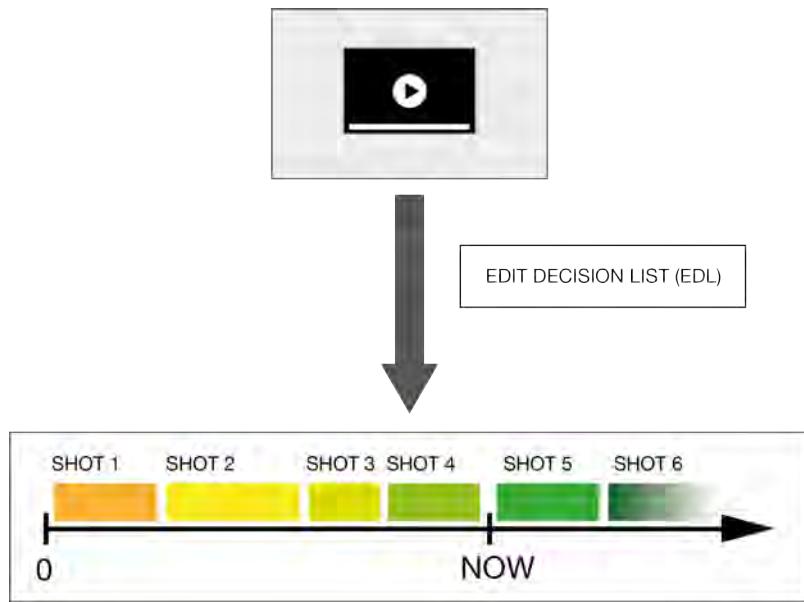


Figure 7.1.3: Illustration of the planned assembly of music video sequences.

In both cases the processing time of the algorithmic editing decisions has to be minimised to reduce waiting times for the viewer. The second method, however, better supports uninterrupted playback because the entire decision-making process is completed in advance, allowing for an improved buffering of the video content.

Building on the decisions about *when* algorithmic decisions are made in the generative editing process, structural editing dynamics then define the editor's creative influence over sequence outcomes. These structural dynamics are based on two different concepts: (1) probabilistic value sets based on selected visual characteristics of the footage used to define a sequence's progression, and (2) linear narrative arrangements, which facilitate the logical progression of story elements.

These two concepts correspond to two essential ideas important to the music video format and editing: audio-visual rhythm and narrative. Rhythm will be defined by the above-mentioned probabilistic structures, which largely determine a music video's look and feel. Considerations towards narrative, in turn, are based on supporting these structural definitions in videos that require the depiction of a more detailed sequence of events. In this context, the different narrative levels described in detail in Chapter 5.4.4, which elaborates on the varying requirements for the rigidity of narrative structural definitions, play an important role.

After deciding on the point in time the shot decisions are made by the software, on the type of footage as well as the order of shots in the sequence, it was then defined how the footage is cut

by *Dividation*. As mentioned earlier, two essential methods were designed for this purpose. The following sections will describe the development and implications of these methods on the generative editing process in more detail.

7.2 Generative Editing Development

The development of *Dividation* produced three distinct versions. *Version one* comprised first explorations into generative editing methods by remixing an existing music video, providing some basic insights and methods for further research. With *version two*, generative editing was first explored using original music videos extending on basic insights gained during the development of *version one*. During this stage, the production process of generative editing, from planning to final editing, was defined as a whole. *Version three* continued to build on the insights gained in the previous experiments by further refining and optimising these methods.

The following subchapters will discuss the experiments chronologically, describing both the challenges and methods developed to derive a generative editing algorithm which facilitates an intuitive creation process and supports the editor's creative practice.

7.2.1 Version One

The first attempts at algorithmic editing explored the utilisation of Markov chains to confirm their applicability for generative editing. Therefore, the initial experiments were designed to provide information on how parameters for music video editing must be defined and on the necessary relationships between them in order to form the basic conceptual structure of the editing algorithm.

Version one of *Dividation* consisted of various experiments to remix an existing music video, *Big Jet Plane* by Angus & Julia Stone (cp. Chapter 5.2.3).

This first stage of generative editing experiments focused on:

- Exploring editing methods based on Markov chains,
- defining key concepts as a basis for further experiments, and
- defining structural models to recreate the music video's original progression, look, and feel.

The first step in this process was a detailed analysis of the music video *Big Jet Plane*, described in Chapter 6.2.3. Based on the analysis, a provisional set of video parameters was selected.

7.2.1.1 Parameter description

For the analysis, the *Big Jet Plane* music video was cut into a collection of individual shots.

During this process, prominent visual features were documented, eventually leading to the definition of four structural parameters:

- shot duration,
- field size,
- the protagonist's location, and
- object shots, which do not depict the protagonist.

Each parameter was chosen to fulfil a particular purpose in the probabilistic description of the editing dynamics. Both, shot duration and field size, served to recreate the music video's visual rhythm. The field size influences the visual rhythm according to the visual representation of content in a shot. Duration, in turn, affects this rhythm by defining the sequence's temporal structure. The other two parameters, which focus on what and who is seen in a shot, define the narrative or logical progression. However, the dominant parameter for describing the video's logical structure is the protagonist's location. This parameter is used to describe a rough sequence of events. The music video *Big Jet Plane* revolves around the idea of being stuck in one place, but wanting to be somewhere else instead. The simple progression of the three different locations in this video therefore constitutes the main aspect of this simple narrative.

Varying levels of detail in the description of these parameters were explored during the first analyses and generative editing experiments. For the first experiment, the following specifications were used.

The duration parameter was implemented by forming three clusters of duration lengths (in seconds). The field size was described by adopting common classifications introduced, for instance, by Ascher and Pincus (2007) based on five distinct qualities that appear during the

video (Figure 7.2.1). Finally, the location parameter and the object parameter were described using three qualities (Figure 7.2.2) and two qualities (Figure 7.2.3), respectively.

T1 = 1 seconds to 4 seconds

T2 = 5 seconds to 10 seconds

T3 = 11 seconds or more



Figure 7.2.1: Examples of field sizes in the Big Jet Plane music video (NettwerkMusic, “Angus and Julia Stone - Big Jet Plane [Official Music Video]”, <https://www.youtube.com/watch?v=yFTvbcNhEgc>).

ECU = Extreme Close Up

CU = Close Up

MCU = Medium Close Up

MS = Medium Shot

W = Wide Shot



Figure 7.2.2: Locations shown in the Big Jet Plane music video (NettwerkMusic, “Angus and Julia Stone – Big Jet Plane [Official Music Video]”, <https://www.youtube.com/watch?v=yFTvbcNhEgc>).

H = Home

O = Outdoors

S = Shop



Figure 7.2.3: Object shots in the *Big Jet Plane* video (NettwerkMusic, “Angus and Julia Stone - Big Jet Plane [Official Music Video]”, <https://www.youtube.com/watch?v=yFTvbcNhEgc>).

O = Object

OC = Object close

For the successful implementation of these parameters in the generative editing algorithm, it was necessary to consider their differing functionality with regard to defining the editing dynamics. Some of these parameters describe visual features, others focus on sequence structure. While the field size and location are correlated, the object parameter had to be treated individually because the object parameter was defined to be applicable only to shots shown in the second segment, which predominantly features objects in the protagonist's room. The field size parameter, here, can be used to describe how often the protagonist is seen. Therefore, this parameter is not applicable to the object parameter. This is why the object parameter was defined consisting of two qualities, including its own description of the field size (e.g., close and not close). The duration parameter, however, applied to all shots.

In this experiment, the duration parameter was directly linked to individual shots because the video only provides a limited amount of footage, i.e., one manually edited sequence. Therefore, the potential for experimentation with shot duration in *Dividation version one* was necessarily limited.

7.2.1.2 Parameter Combination

After translation basic video characteristics into parameters for the algorithmic process, it was necessary to develop and test different decision models combining and placing these parameters into relationships with each other to form the basis of the generative editing process, building on the concept of a Markov chain.

First, the original sequence was described according to a comprehensive documentation of all shot transitions. Figure 7.2.4 shows one complex combinatorial structure resulting from a description method based on the three visual parameters. This model clearly illustrates the need for structuring methods that are simple, intuitive and efficient in order to experiment with varying numbers of parameters.

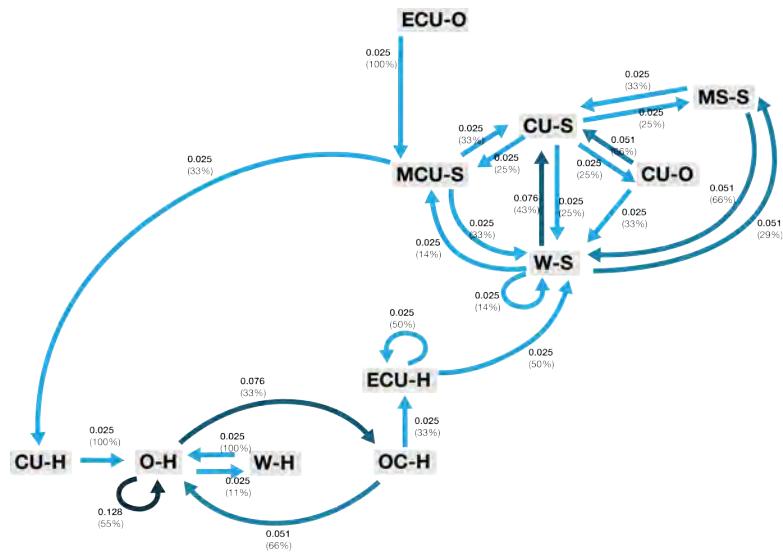


Figure 7.2.4: Combined parameter model created using the three visual parameters.

Example 1.1¹²⁴ shows three generative videos created using a simple algorithmic implementation of this model. Figure 7.2.5 presents a pseudo code of an algorithmic shot decision based on the ideas and structures of this model.

```
If this is the first shot
    Play first shot
Else
    Read previous state
    Find state probabilities for next shot
    Determine next state with random number according to probabilities
    Form a set by fetching all shots that match the new state
    If the set is 1
        Play this shot
```

¹²⁴ Julia Stefan. "Dividation Version 1 - Big Jet Plane, combined structural definitions", YouTube video, 4:05, 9 February 2015, <https://www.youtube.com/watch?v=QppqZvsWWdU>.

```
Else, if the set is 0 and decisionCount is smaller than 2
```

```
    Find new state
```

```
Else, if the set is larger than 1
```

```
    Get random shot from the set
```

Figure 7.2.5: Shot decision algorithm describing the shot decision process.

Because the music video remix algorithm has to operate with a small number of available shots taken from the original sequence, the logic behind the algorithmic decision process in this first experiment remains rather limited. In order to increase variability in this experiment, shots could be repeated within the sequence, and audio-visual synchronisation was not considered yet.

The results of the experiment underscored the challenge of creating the perception of sensible decision-making. While the above structure includes a large amount of detail, it reduces the amount of control and may lead to a largely homogeneous set of sequence outcomes due to the lack of variability in the probabilistic structures used.

The structure used here features eleven individual states. Each available shot corresponds to one of these states, e.g., an extreme close-up outside or a medium close-up in the shop. By adding more parameters to this structure, the number of states may increase exponentially, leading to a complicated, rigid base structure, which may become increasingly difficult to implement.

Example 1 further highlights the issue of control, or rather the lack thereof. Once the initial probabilities have been defined, the author can no longer influence the sequence. This means that the algorithm will only make decisions according to the above model, which describes shot transition probabilities over the entire sequence. In some cases this may yield coherent and unexpected results, particularly when using a relatively open probabilistic structure. In the above example, the sequence outcomes show little variability because only one set of probability values was used for the entire sequence and due to the way these probability values were defined.

One of the primary goals defined for this first experiment was the achievement of similarity to the original music video. Therefore, it was necessary to find a method which could imitate this video's editing dynamics more accurately. Because music video sequences may display

dramatic differences in editing dynamics according to the video's narrative state and sonic progression, various means of temporal control were introduced, which will be described in the following section.

7.2.1.3 Segmentation

One way to simplify this complex network was to correlate the documented editing dynamics to the video's segments, rather than considering the video as a whole.

In the case of *Big Jet Plane*, as mentioned before, three visually distinct segments are established in the original sequence through the video's narrative and visual progression. Figure 7.2.6 shows a model of transition networks similar to Figure 7.2.4 but split into these three segments.

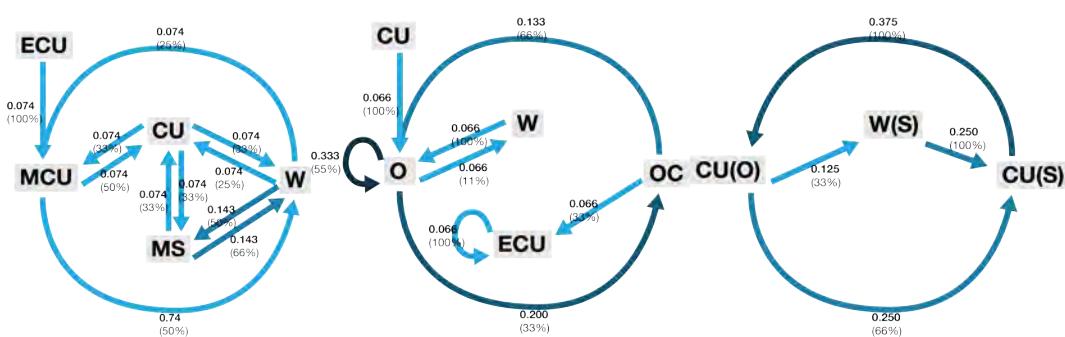


Figure 7.2.6: Segmented transition model split into three sections.

The graphs only depict those parameters that change within a given segment. Because the location is a dominant identifier in the definition of these segments, it is not explicitly included in the first two graphs. After the first shot, which is also the only extreme close-up in the sequence, segment one only takes place in the shop. The second segment shows the protagonist's home. Because the third segment switches between the shop and the outdoors, the location parameter had to be considered and is indicated in brackets.

The implementation of segmentation as a simple temporal identifier introduces an element of logical control, which tentatively renders the generative editing process more intuitive for the author and allows for greater accuracy towards his or her conceptual intentions. The process implied in the previous model, which depicted all shot transitions in one single graph, does not include temporal checkpoints and may therefore be too autonomous and result in incoherent results. The segmented model provides more control over shown footage at

particular points throughout the video by limiting access only to certain kinds of footage depending on the sequence's temporal state.

The algorithmic implementation of a single decision using this model is similar to the one described in the previous section. However, the temporal division results in a more specialised progression. Example 1.2¹²⁵ shows some results generated with the help of this model.

By visualising sequences as shown in Figure 7.2.6, we may gain a better overview of the music video's conceptual editing structure, which may facilitate a more intuitive creation process.

The model presented here, however, still constitutes a highly conceptual attempt at formulating the basis behind generative decision processes. Each parameter is based on different ideas and individual structural applications, and would have to be implemented with precise functions that are tailored to the very individual ways in which these parameters influence the sequence. In their current state, as described so far, they were too heterogeneous for further experimental stages where the use of different numbers and types of parameters will be examined.

7.2.1.4 Abstraction

Another step in optimising the algorithmic efficiency and balancing the level of complexity when describing shot characteristics and dynamics was to modify the parameters towards a common design. In order to support modularity, parameters and filtering processes had to be standardised by using the same design for each parameter. This led to the limitation of each parameter to a maximum of three qualities. The object parameter was removed. Shots that were previously described with this parameter were annotated to fit into the new model described below.

The three parameters used in this model consisted of simplified descriptions of location, field size and shot duration. Furthermore, the parameters were now treated individually, rather than in a combined structure.

¹²⁵ Julia Stefan. "Dividation Version 1 - Big Jet Plane, segmented structure", *YouTube* video, 4:05, 9 February 2015, <https://www.youtube.com/watch?v=PGH-UDbfsU0>.

Figure 7.2.7 shows an unsegmented example of the three parameters used in this revised model.

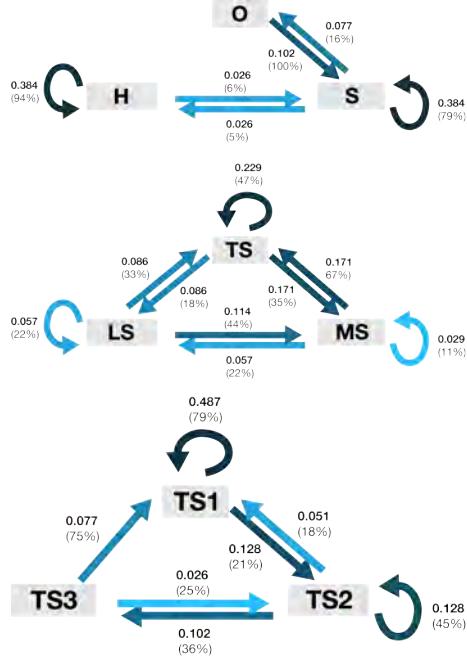


Figure 7.2.7: Shot transitions according to the individual parameters with reduced complexity.

This approach facilitates the implementation of a larger number of parameters by increasing the algorithm's modularity.

Example 1.3¹²⁶ shows video sequences created using this abstract parameter model. The decision process for a single shot is described in Figure 7.2.8.

If this is the first shot Play first shot Else For each parameter Read previous state Find state probabilities for next shot Determine next state with random number according to probabilities Form a set by fetching all shots that match the new state Intersect three sets into final set

¹²⁶ Julia Stefan. “Dividation version 1 - Big Jet Plane, segmented abstract structure”, YouTube video, 4:05, 24 February 2015, <https://www.youtube.com/watch?v=6oz9kEfOYbw>.

```

If the final set is 1
    Play this shot
Else, if the final set is larger than 1
    Get random shot from the set
Else, if the final set is 0
    Repeat decision process
Else, if the decision process was unsuccessful for 20 times
    Choose random shot

```

Figure 7.2.8: Shot decision algorithm for a single shot based on the abstract parameter model.

A principal advantage of this design is the standardised implementation of the parameter model, which helps to both simplify the code and provide a streamlined, intuitive creation process. The pre-defined parameter structure becomes the author's tool. It provides a closed off space in which conceptual creativity can unfold. Although some of the author's creative freedom is removed in the process, this method also serves to preserve and redirect the author's efforts towards creative goals, in addition to devising the technical background of a generative video's technical structure.

In order to derive a system for generative editing, we must impose some limitations on the user of this system. Candy and Edmonds (2002, 25) already pointed out that a "simple system", therefore, "may not have every conceivable function available within it, but for creative purposes, the constraints that it imposes may be a positive stimulus to new ideas".

Example 1.3 reveals several challenges to the design. Segmentation was applied in a similar fashion to the previous example.

The algorithmic decision process based on this model may become less predictable because an individual decision is made for each parameter, without a direct relationship between them. This introduces the possibility of parameter combinations which did not exist in the original footage, and requires further definitions to facilitate an appropriate decision making process. Without these definitions, the algorithm may require more iterations until it results in a suitable decision. In fact, a simple escape condition was introduced which counted the number of times the decision process has been repeated, choosing a random shot if this number exceeds a certain amount.

7.2.1.5 Discussion

The development of *version one* was declared concluded when the generative outcomes yielded satisfactory results, i.e., when they showed visual accuracy towards the original music video's structure. However, there was no attempt to fully recreate the video since the design of *Dividation* is not orientated towards finding one correct solution, but to create a system that allows for variability in sequence generation. The development of *version one* delivered the basic processes, which would then help to create an original generative music video specifically produced for this purpose.

The experiments conducted in connection with *version one* tested the use of text files, XML and finally MySQL databases to provide shot information to the editing scripts. Sequence creation was done in real-time entirely within JavaScript.

With *Dividation version one*, an existing music video was analysed and remixed to examine the video's editing dynamics, and to devise a simple method for an algorithmic recreation of this video. A simple algorithm was designed based on an iterative decision-process using Markov chains to select each shot in the music video sequence. Figure 7.2.9 illustrates the shot selection algorithm.

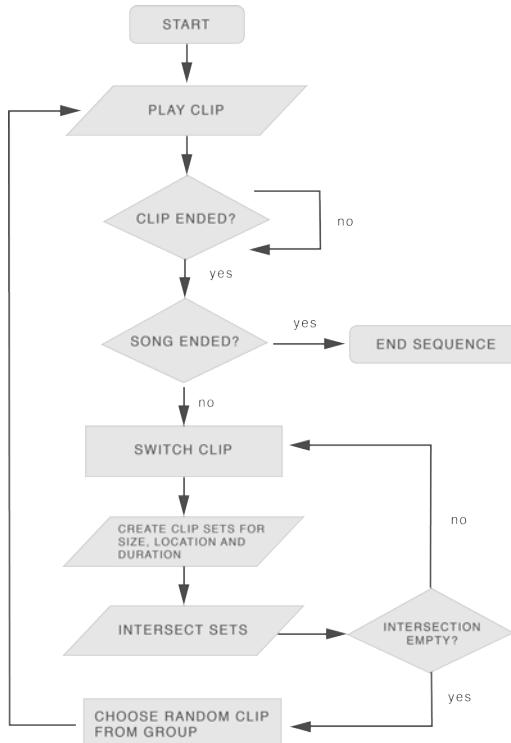


Figure 7.2.9: Algorithm for a single shot decision for *Dividation* version one.

First, a shot is chosen and played. The first shot in the sequence has been predefined. In accordance with the Markov approach, each following shot depends on the visual characteristics and the duration of the preceding shot, as well as a probabilistic structure of shot transitions derived from the original sequence. Every time a shot is played, a decision is made for the next shot to be played. Figure 7.2.9 shows a simple visualisation in which this decision-process is repeated until a suitable set of footage is found. This reiterative decision-process terminates once the song is over.

The guiding principles in this first version of the generative music video editing process were devised by (1) analysing an existing sequence for a shot transition model, (2) developing and testing parameter models, and (3) combining and using parameters to form a decision-process.

Version one applied a process in which decisions were made in real-time during playback. While the decision process for this set of experiments is comparatively simple due to the limited amount of footage available, it was foreseeable that the increasing complexity of the following experiments would require a process that could ensure stable playback. *Big Jet Plane* uses shots of very long durations, but many music videos tend to display a much faster editing rhythm where many shots may only consist of two or three frames, with a length of approximately a tenth of a second. The amount of footage handled in the generative decision process, as well as the complexity of the implemented editing dynamics, was expected to lead to difficulties in following experiments.

Version one opened up several other areas of inquiry. After testing basic methods for generative editing, the next phase of experiments was designed to comprise original video material and to illustrate a full generative editing process from the initial planning, to production, editing and post production and, finally, to sequence generation. For this purpose, it was necessary to review the parameter design and to develop methods of establishing relationships between these parameters. *Version one* introduced an abstracted parameter model to standardise the parameter design. With the next phase of experiments, the development of both parameter models based on this design and increasingly modular editing algorithms was continued. These experiments further investigated what Bizzocchi (2011, np) referred to as “the right balance between variability … on the one hand, and aesthetic control on the other”.

7.2.2 Audience Survey

A survey was conducted to receive a first feedback concerning the effectiveness of the algorithms designed for *Dividation version one* (cp. Appendix B).

For the survey, two sets of five one-minute videos were displayed to the participants. Each set featured an existing popular music video, which was shown in

- its original composition,
- in two versions created using algorithmic decision-making, and
- in two randomly assembled versions.

The first set consisted of video samples assembled from the music video *Gangnam Style* by Psy (cp. Chapter 5.2.1). For the second set, the music video *Big Jet Plane* by Angus and Julia Stone was used (cp. Chapter 5.2.3), with the sequence generation based on *version one* of *Dividation*.

Both videos were created using the methods described in the previous sections with regard to sequence analysis and generation.

Randomised elements were combined with a minimal amount of control over clip positioning to ensure that footage featuring the musicians' performance of the song was in synch with the music.

Participants were asked to complete rating scales for each video thereby agreeing or disagreeing with the following statements:

- I enjoyed watching this video.
- The video was engaging.
- The video made sense.
- The video felt rhythmic and suited the music.
- The video was flowing and non-disruptive.

Additionally, participants were free to name videos they particularly liked or disliked and to add further comments regarding their choices and tastes.

7.2.2.1 Results for Gangnam Style

For the first set of music videos featuring *Gangnam Style*, 93 sets of data were collected and analysed. The majority of participants were female (cp. Figure 7.2.10) and aged between 21 and 40 (cp. Figure 7.2.11). Most participants stated that they were watching music videos on a weekly basis (cp. Figure 7.2.12).

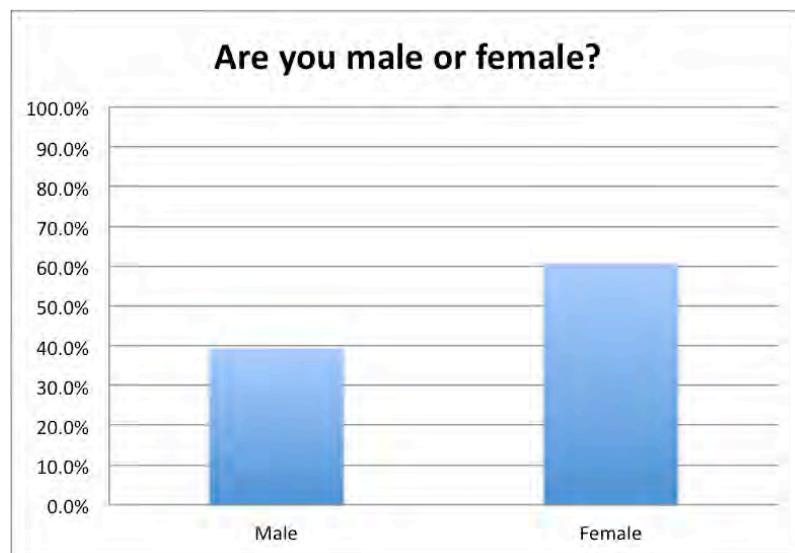


Figure 7.2.10: Percentage of participants by gender for *Gangnam Style*.

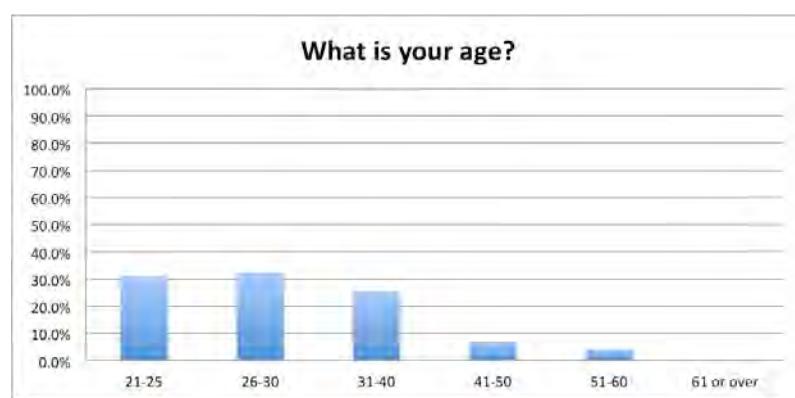


Figure 7.2.11: Percentage of participants by age for *Gangnam Style*.

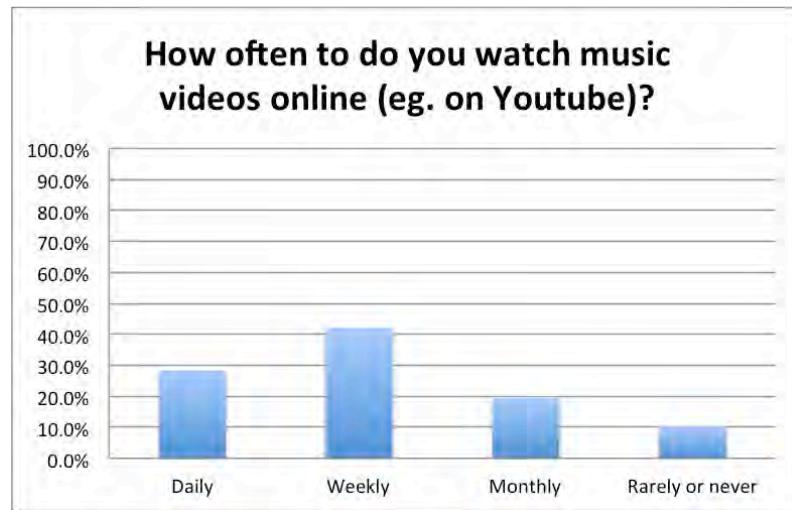


Figure 7.2.12: Participants' viewing habits for *Gangnam Style*.

The ratings for *Gangnam Style* indicate a general preference for the original version (O), as well as for one of the algorithmic compositions (A1). The overall results (cp. Figure 7.2.13) indicate that the participants preferred composition A1, ranking it highest in the categories “I enjoyed watching this video” and “The video is engaging”. However, the original composition was rated best in all other categories.

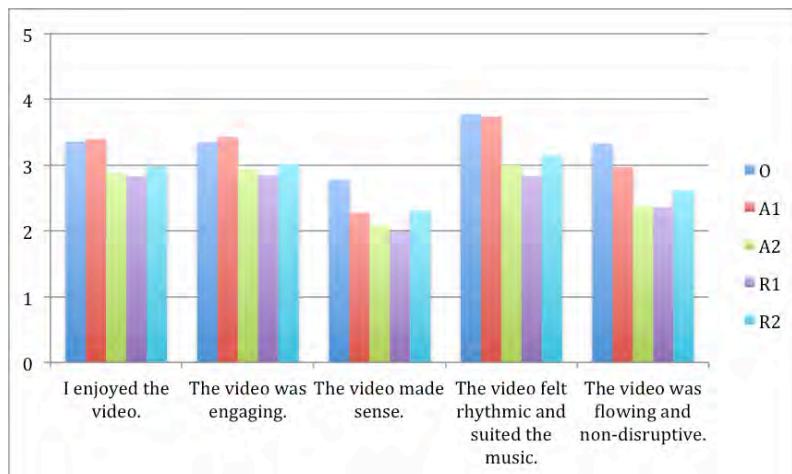


Figure 7.2.13: Survey summary for *Gangnam Style*.

Similarly, the participants mentioned the composition O and A1 as videos they liked (cp. Figure 7.2.14). However, when asked which videos they disliked, composition O was chosen more often than A1 (cp. Figure 7.2.15). Although the results indicate that the original composition was generally perceived to be more rhythmic and more in sync with the music, several participants mentioned that the use of slow-motion footage in composition O was perceived as disruptive and inappropriate.

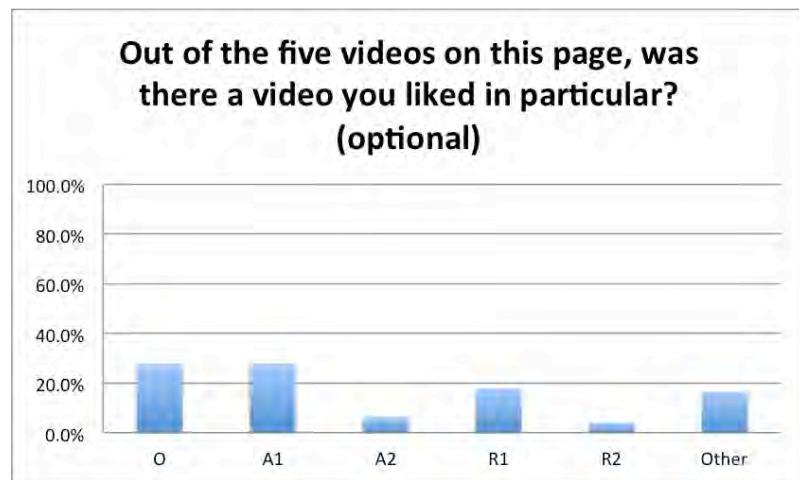


Figure 7.2.14: Results for *Gangnam Style*: Videos "liked" by the participants (in percent).

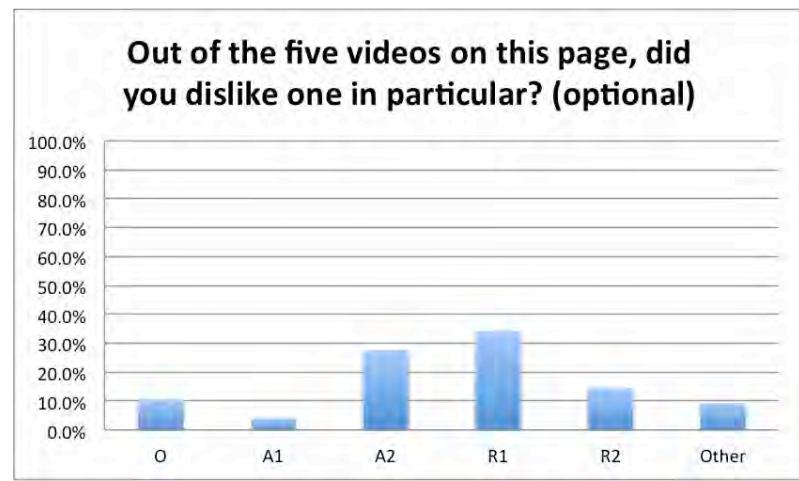


Figure 7.2.15: Results for *Gangnam Style*: Videos "disliked" by the participants (in percent).

The two random compositions (R1 and R2) were met with mixed views. Some comments implied that the use of fast-paced, random-seeming material is pleasant to watch when these characteristics are applied methodically and consistently.

Results across different target groups (age, viewing habits) were consistent.

7.2.2.2 Results for *Big Jet Plane*

For the second set of videos featuring *Big Jet Plane*, 76 complete datasets were collected and analysed. Again, the majority of participants were female (cp. Figure 7.2.16) and aged between 21 and 40 (see Figure 7.2.17). Most participants stated that they were watching music videos on a weekly basis, similarly to the previous sample group (cp. Figure 7.2.18).

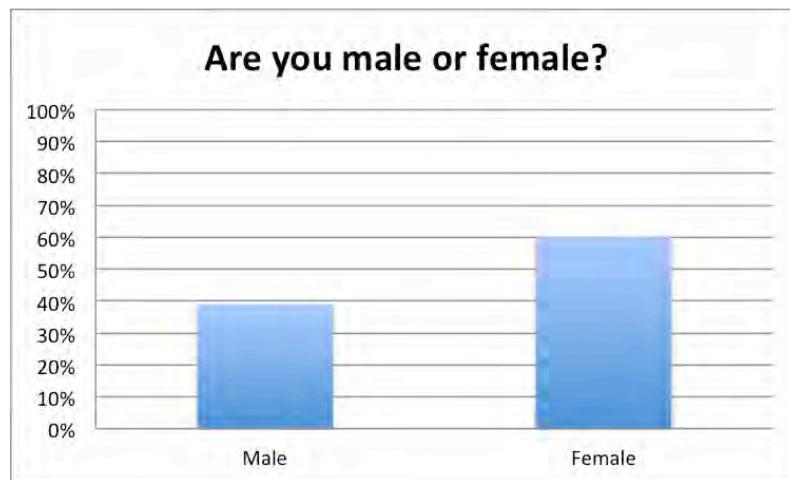


Figure 7.2.16: Percentage of participants by gender for *Big Jet Plane*.

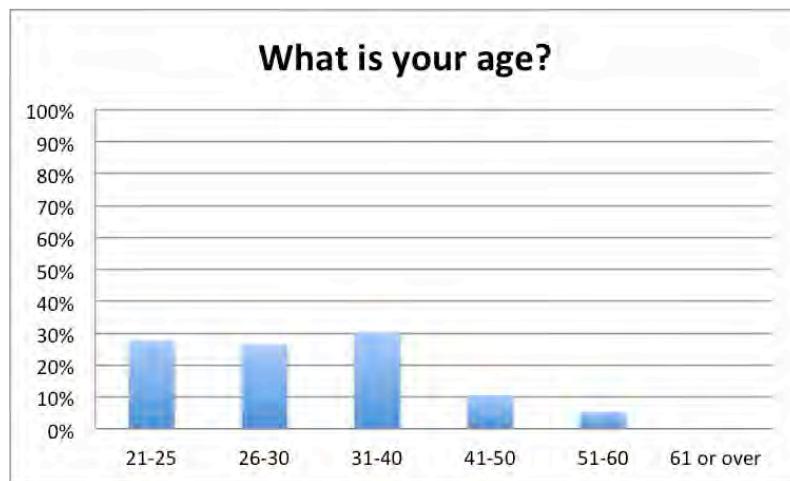


Figure 7.2.17: Percentage of participants by age for *Big Jet Plane*.

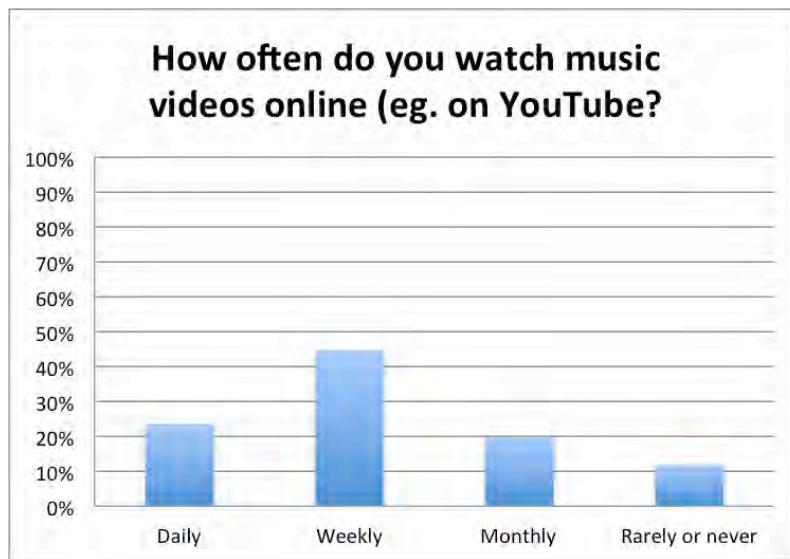


Figure 7.2.18: Participant's viewing habits for *Big Jet Plane*.

Compared to the results for *Gangnam Style*, which showed relatively consistent trends, the results for *Big Jet Plane* are more ambiguous and rating values are less distinguishable. The ratings for *Big Jet Plane* indicate a general preference for the original composition.

Composition R1 was rated the most “flowing”, but was also identified as the most disliked. A1 and R2 received the lowest ratings (cp. Figure 7.2.19). R2 was also the most liked video (Figure 7.2.20), as well as the second most disliked video in the set (Figure 7.2.21).

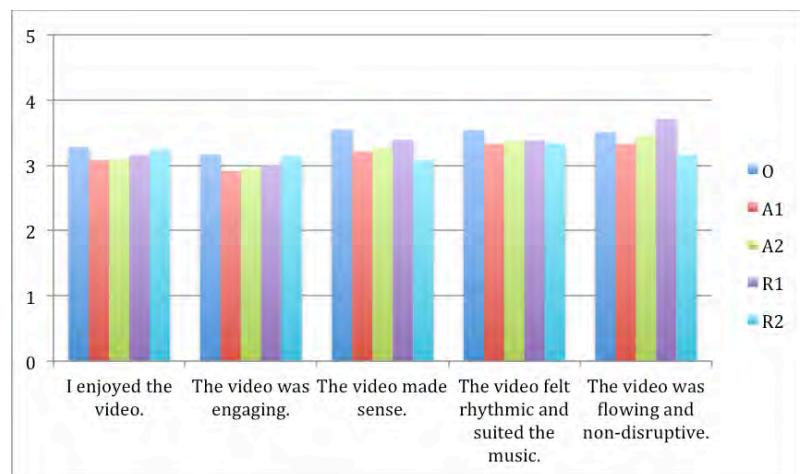


Figure 7.2.19: Survey summary for *Big Jet Plane*.

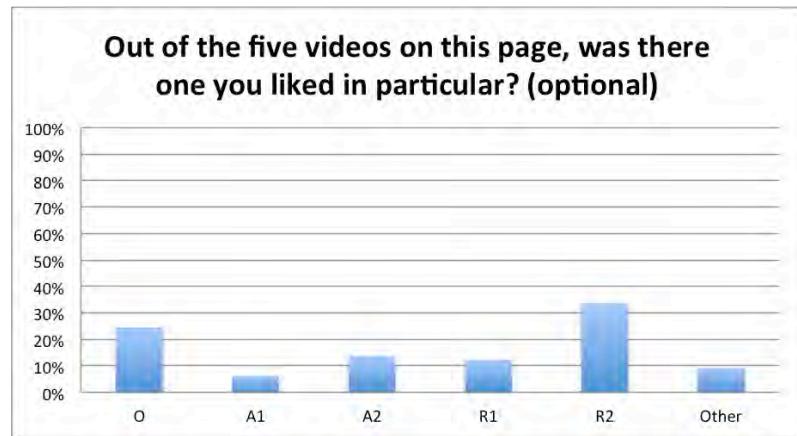


Figure 7.2.20: Results for *Big Jet Plane*: Videos "liked" by the participants (in percent).

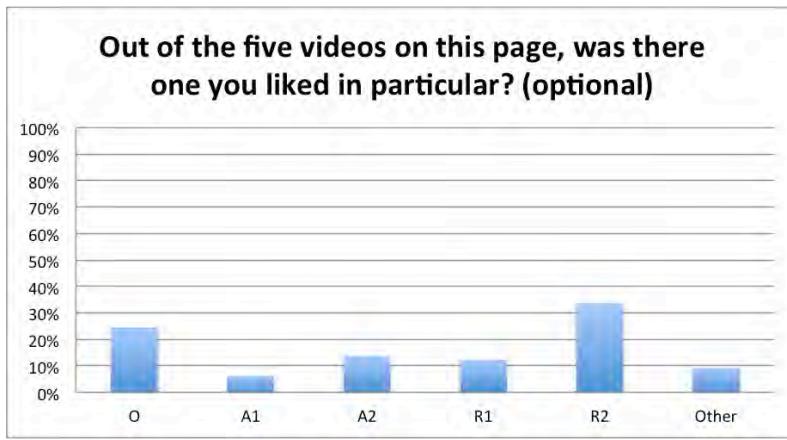


Figure 7.2.21: Results for Big Jet Plane: Videos "disliked" by the participants (in percent).

Some variation in the preferences was found among different target groups, but few distinguishable trends emerged.

By differentiating the results according to the participants' age groups, the group of 21-25 year olds showed some differences compared to other groups. These participants preferred randomised sequences over algorithmically edited sequences.

Weekly viewers preferred the original and the randomly edited sequences. These participants, however, also rated one of the random sequences as both the most liked and the most disliked video in the set. Monthly viewers, in comparison, preferred the original composition and disliked most other sequences, except for one of the algorithmically edited sequences. The results obtained for viewers who indicated that they watch music videos on either a daily basis or rarely did not display distinguishable trends.

In general, the algorithmic compositions ranked average to low in this set. Of particular interest was the sequence R2, which received both very high and very low ratings. The comments received suggested that some participants, especially those who did not like the overall theme of the video, preferred the more chaotic and randomised order of shots in composition R2. Another aspect in this sequence preferred by participants was the visual content in some of the shots that could not be seen in other versions, and therefore differentiated the sequence from others. In addition, although a majority of participants stated that they prefer a clear story, various comments suggested that participants enjoyed the variation of visual content and sometimes even favoured a particular version because it was different from what they had seen previously.

The algorithm's preference of visual and rhythmic editing over narrative coherence might be one reason why the algorithmic versions of this video were ranked comparatively low. This way, the program imitated the original video's story and style, but the settings were not precise enough to sufficiently recreate the storyline. Although editing in A1 and A2 turned out relatively similar to the original video, several comments described those sequences as "boring" and "random". Participants who disliked the algorithmic sequences seemed to perceive the random sequences more informative because of the variety of content shown in those sequences.

7.2.2.3 Discussion

The results were affected by the survey's viewing context, which differed from a common music video viewing experience. First, viewers were asked to watch a section of a music video a total of five times. Because the music video remixes depended upon existing, manually edited video sequences with a limited amount of footage available for re-editing, recombination possibilities for these newly created sequences were limited. Furthermore, the two sets of music videos depended on vastly different original material. The first music video features an energetic and humorous sequence based on a fast-paced pop song, whereas the second consists of a simple narrative for a slow indie folk song. The stylistic and atmospheric contrast between the two sets of videos may therefore have influenced the participants' perception.

With these potential issues in mind, the survey was designed to reveal some of the advantages and disadvantages of the use of generative music video editing, and to learn more about what viewers may be looking for.

The literature contains very few studies on how viewers actually perceive music videos. Further studies may therefore have to investigate the importance of what we might understand as the quality of editing and coherent progression of a music video, in comparison to its overall visual content, e.g., images that evoke emotions, and images that showcase the artist. The *Big Jet Plane* music video set has revealed that some viewers seem to prefer sequences, which depict the musician and provide a variety of shots, including images that may bear a greater emotional affect.

The brief investigation into the popularity of the *Gangnam Style* music video in Chapter 5.2.1 suggested that audiences liked the video because of its humorous content. The most important aspect in the editing of this music video sequence may then lie in the highlighting of both the humorous scenes and memorable moments in the song. This is a task beyond the capabilities of the automated processes explored in this project and thus emphasises the role of human editing and prioritisation as part of the generative video workflow.

Differing viewer preferences were also a confounding element in this survey. For instance, while some viewers may rate music videos negatively because they are perceived as disjunctive and illogical, other viewers may enjoy this particular aesthetic style. Verbal participant feedback suggested that two distinct groups of viewers can be distinguished; (1) viewers who prefer to see a consistent, or classical, linear storyline, and (2) viewers who enjoy seeing a wide variety of material to make sense of the whole scenario in a less linear fashion. It is possible that this second group is more accustomed to what Vernallis (2007, 130) referred to as a typical function in music video, requiring viewers to “extrapolate from the shots” in order “to give meaning to a juxtaposition”. This also relates to a description by Goodwin (1992, 70), who stated that “visual pleasure is present not so much at the level of narrative, but in the making musical of the television image”.

Goodwin further elaborated that disjunctive and random-seeming images are an essential aspect of music video, designed to limit the viewer’s immediate understanding of the sequence, and therefore motivate repeated viewing. In this context, Vernallis (2001) also pointed out that the editing in a music video will commonly make itself noticeable. This serves to prevent “powerful images from acquiring too much weight and stopping the flow of information. The editing thus preserves the video’s momentum and keeps us in the present” (Vernallis 2001, 22).

Some participants stated that their preferences depended on the balance of audio-visual rhythm and narrative elements. Confirming previous descriptions of music video aesthetics and dynamics discussed in Chapter 4, the music video, therefore, should establish its own logic, but should also maintain a clear connection to the song.

The algorithmically assembled sequences seemed to function comparatively well for the *Gangnam Style* set. This music video is inherently perceived as being non-narrative. The algorithmic sequences managed to create greater coherence for some viewers through

persistent editing dynamics, which highlighted the rhythmic elements of the music video over the simple narrative section shown in the original. The algorithmic sequences were less successful in the case of *Big Jet Plane*. As already mentioned, this music video is driven by a simple narrative, which unfolds in a slow progression. Here, the algorithmic sequences did not accurately recreate the series of events shown in the original, but they also did not display the kind of consistency arising from one of the randomised sequences, which juxtaposed a wide variety of shots, and therefore managed to give an impression of coherence.

Some comments indicated that, apart from preferences in progression, participants also preferred certain videos because of their visual content and videos that differed from other versions in the set, thereby providing added value in making sense of what is seen in the context of repeated viewing.

7.2.3 Version Two

At the heart of the development of *Dividation version two* was the goal to define a generative editing process from start to finish, using original video material created specifically for this purpose.

The following set of experiments was performed using two different original music videos, i.e., *Majestic* by the Brisbane band The Sabre Siren, and *Look... The Sun is Rising* created as an entry in a music video contest for the band The Flaming Lips.

As a next step in the development of *Dividation, version two* aimed to

- define the generative editing process for an original work and complete workflow,
- develop methods for audio-visual synchronisation, and
- refine the editing algorithm with a focus on extending its modularity.

In addition to defining the editing of a complete work, methods of control were further examined in this stage. Because of the simple approach in *version one*, which was based on recreating an existing music video with a limited amount of footage, the aspect of creative control so far had not received much attention. The development of *version two* therefore endeavoured to find an efficient process ensuring an accurate decision making through the definition of parameter relationships, decision constraints, and methods of altering the

probabilistic decision process, in addition to maintaining the concept of segmentation introduced with *version one*.

The following sections will describe the two music videos used in these experiments as well as their role in the development of *Dividation version two*.

7.2.3.1 Majestic

Majestic was shot in a single, largely empty studio space during the course of one day (Figure 7.2.22). Inspired by the song, which addresses the artificiality of mass media, the video was planned to feature a conceptual narrative (cp. Chapter 5.4.4). The simple story therefore consisted of the protagonist, Kate, watching and eventually entering the band's televised performance. The spatial realities of Kate's surroundings and the televised images merge until the show is over and she is alone again. The video shows the band both performing and setting up in order to highlight some of the processes required to produce these images.



Figure 7.2.22: Production of Majestic.

During pre-production a list of shots required for this experimental stage was identified, rather than using a strict storyboard. This list included close-ups, medium shots and wide shots of all protagonists, as well as different combinations of visual elements, for example, Kate alone and Kate performing with the band. Similarly, several takes focused on capturing the band alone during setup and performance, then on capturing the band in front of the

paper silhouettes, and of course on capturing the band with Kate. Figure 7.2.23 shows several shots recorded according to these basic requirements.



Figure 7.2.23: Examples of different types of shots in the *Majestic* music video.

The video's conceptual narrative can be outlined in six steps. The following points are listed in temporal order, but can overlap and occasionally change during the actual video sequence.

- 1) Kate enters the space, sits down in front of the television set and turns it on;
- 2) The band is shown setting up the space for their performance;
- 3) The band performs;
- 4) Kate watches the screen;
- 5) Kate enters the performance space while the band is playing;
- 6) The band disappears and Kate is alone again.

The white studio space was used as a steady element providing some continuity and orientation for the viewer. The studio space was further used, on the one hand, to create a distinct visual style, and on the other hand, to facilitate and simplify the generative editing process. The video's progression was largely based on adding and removing people and visual elements from the space so that the visual features communicated to the generative editing system remained comparatively simple.

7.2.3.2 *Look... The Sun is Rising*

The video for *Look... The Sun is Rising* constituted a minor part of this experiment. It was used, on the one hand, to extend and test the modularity of some of the methods developed during *Dividation version two*, and on the other hand, to test the development of software with openFrameworks, rather than HTML5, JavaScript or PHP.

A very simplistic approach was taken for producing this video. The video exclusively consists of close-ups and extreme close-ups of the protagonist. Its progression is defined by the

increased staggering of effects, changes in field size and increasing tension in the protagonist's expression (Figure 7.2.24).

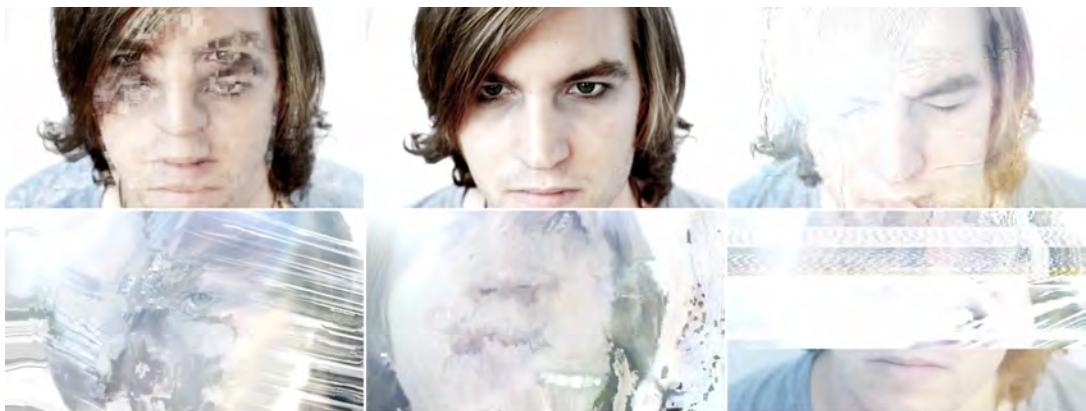


Figure 7.2.24: Examples of the progressive states in the *Look... The Sun is Rising* music video.

The following description of the music video was given by the band:

"I know this is normally an optimistic statement: 'look...the sun is rising!'...but here, it signals the enemy...like the way a vampire dreads the dawn.

There is a time at 4 or 5 or 6 in the morning when, after you've been up all night, there is sense that you have escaped from your life...from your routines...from your responsibilities...but you don't know it; you are just floating unaware of the fake freedom you are experiencing...and then you look over your shoulder and you catch a glimpse of the horrible solar rule – the rule that says, 'a new day will come'...the rule that says, 'time marches on'.

We are, if we are lucky, obedient slaves to the wonderful sun... but, yeah...we are slaves...we owe everything to it...and we know we cannot escape it.

*When the sun comes up, it means the world is awake again and that I must be awake again...awake again to the truths that the sun has taught us: that nothing lasts forever... that everything dies. But love can burn more brightly and more intensely despite knowing this."*¹²⁷

The video sought to illustrate this inner process and the struggle of coming to terms with the feelings described above by focusing solely on the protagonist's expression.

¹²⁷ Genero. "The Flaming Lips - Make the Music Video for Look... The Sun is Rising", Genero. <http://genero.tv/theflaminglips/>.

Similarly to the video for *Majestic*, the video for *Look... The Sun is Rising* was shot in front of a neutral, white background allowing to focus on several simple visual features during the editing process. This way, it was possible to experiment with a number of different juxtapositions while maintaining visual continuity.

The development of generative editing methods for this video was comparatively short and later abandoned in favour of a more extensive development of the *Majestic* video, once insights on further parameter definitions and modular functions could be found. Similarly, the development in openFrameworks was terminated due to its comparative disadvantages for browser playback and online streaming. As an open source toolkit, openFrameworks was designed for the creative use of code with a particular focus on artistic applications. It is particularly suited for live installations, mobile applications and general experimentation with audio-visual media. However, providing an application suitable for stable streaming and playback using openFrameworks would have occupied a large part of the research efforts. The use of HTML5 with JavaScript provided this functionality to a degree that was deemed sufficient for the current research goals. The use of openFrameworks may be considered again in future research projects.

7.2.3.3 Grouping and Synchronisation of Footage

Example 2.1¹²⁸ shows the first interface created for *Dividation version two* using the *Majestic* music video (cp. Figure 7.2.25). Here, different types of footage were defined according to phases in the music video's progression, and shown in six channels. In total, five phases were described and linked to particular sections of the song. Each phase was defined by several actions and visual elements. This process was tested to find a method for structuring footage without analysing a manually edited linear sequence. As a starting point, the approach provided some insight regarding the creation of narrative logic, synchronisation, and automated cutting.

¹²⁸ Julia Stefan. "Dividation Version 2 - Majestic, synchronizing footage", YouTube video, 3:51, 12 February 2015, <https://www.youtube.com/watch?v=Ys0KnYrLDJc>.

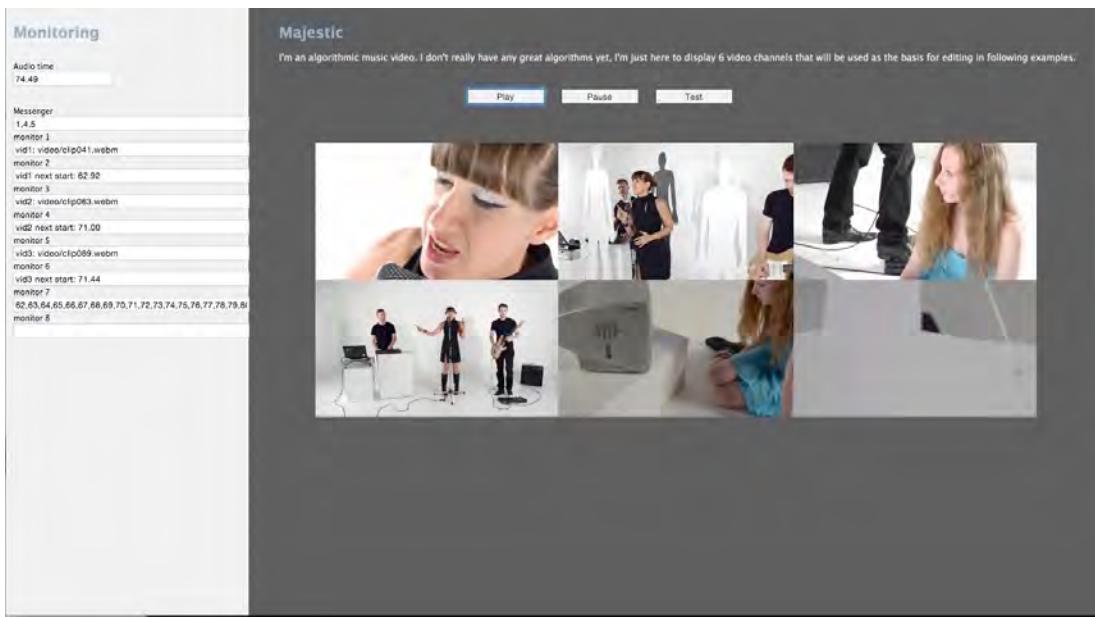


Figure 7.2.25: First experimental interface for *Divination* version two.

Phase 1:

- Kate enters the scene;
- Kate engages with the television set;
- No paper silhouettes;
- No band.

Phase 2:

- The band sets up the performance space;
- No paper silhouettes.

Phase 3:

- The band performs;
- No paper silhouettes.

Phase 4:

- Kate enters the performance space;
- The band sets up the performance space;
- Paper silhouettes are visible.

Phase 5:

- The band performs;
- The paper silhouettes are visible;
- Kate can be seen in the performance space.

The first challenge which had to be solved was audio-visual synchronisation. Compared to *Big Jet Plane*, for *Majestic*, musical performance was regarded a much more dominant aspect.

To achieve an accurate synchronisation, each clip that needed to be synchronised was recorded with a time code indicating its position in the song in order to determine when this particular clip may start and stop during playback. Thus, the song became the driving element to which temporal events are linked.

The interface depicted in Figure 7.2.25 shows synchronised footage in the top three channels. Example 2.1 therefore gives an overview on how much, and what kind of synchronised footage is available at given times throughout the video. This way, we can identify which sections of the music video may have a stronger focus on performance, and which sections will have to foreground narrative elements. The bottom left channel shows one continuous performance of the song, whereas the other two channels show unsynchronised footage. These clips were chosen randomly according to their assigned phases.

In a next step, a simple program combined these channels into a single sequence by cycling through and displaying each channel for five seconds, when possible. Example 2.2¹²⁹ shows the result of this process.

¹²⁹ Julia Stefan. “Dividation Version 2 – Majestic, first combination into a sequence”, YouTube video, 3:06, 12 February 2015, <https://www.youtube.com/watch?v=qRuS69i2N38>.

7.2.3.4 Defining the Editing Dynamics

After solving requirements for audio-visual synchronisation and basic shot categorisation, the first creative concern during the development of generative editing methods for original works must be the definition of detailed editing dynamics. With *Version one* an existing music video was remixed, and it therefore utilised a summary of observed editing dynamics from the original sequence.

Using *Majestic*, *version two* first explored video structures defined by phases, but following experiments eventually had to return to the processes defined in *version one*. A large part of the generative music video created for *Majestic* was therefore developed using dynamics derived from manually edited linear video sequences of the original *Majestic* music video. While the experiments in *version one* established a simplified and standardised parameter model, the first stage in the development of a generative music video for *Majestic* adopted some of the analytical processes used at the very beginning to inform the definition of essential visual characteristics. Figure 7.2.26 shows a selection of notes illustrating this conceptual process.

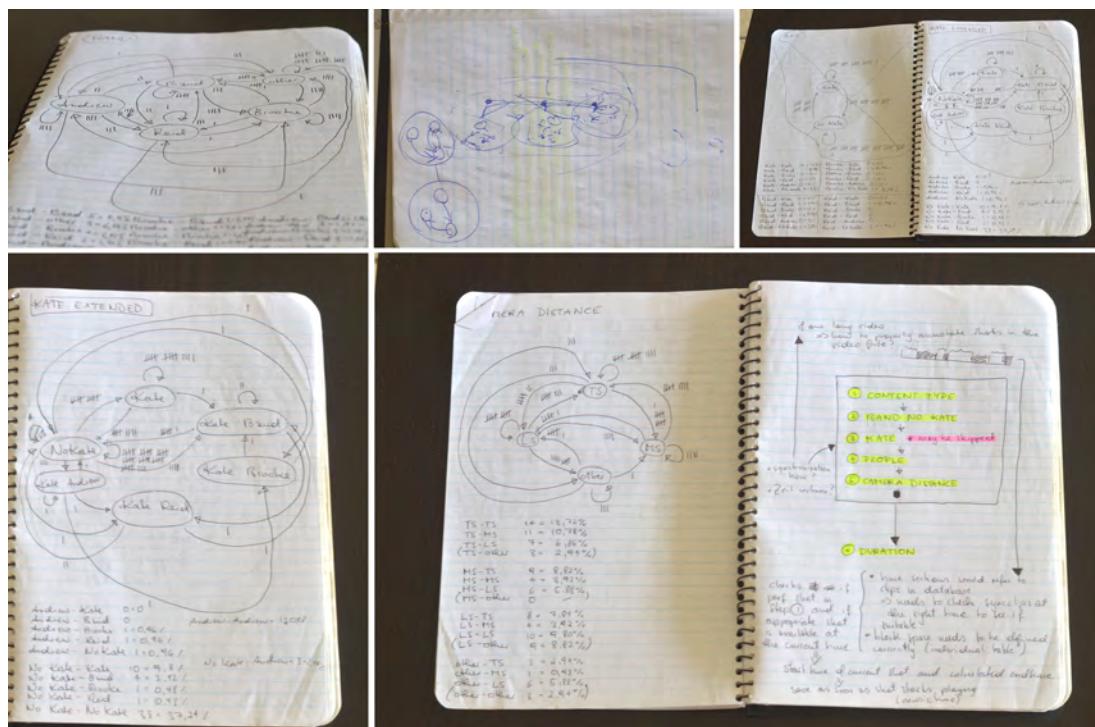


Figure 7.2.26: Conceptual notes written at the beginning of *Divination* version two.

A number of different structures were tested to describe the occurrence of the band and its individual members, the protagonist Kate, and other visual parameters such as camera

distance. Eventually, the experiments showed that the parameter model used for *Big Jet Plane* could not be applied to the video for *Majestic*. When describing the occurrence of the band, for instance, using the previously designed three-way logic, it was possible to apply the parameter to each individual member. But the question arose, how this parameter then could be used to describe the occurrence of two or all band members? Similarly, it became difficult to group the occurrence of the paper silhouettes with other visual characteristics and maintain a consistent intuitive logic in the parameter design.

Similar experiments were undertaken using the music video for *Look... The Sun is Rising*. Because of its narrative structure, outlined earlier in the chapter, this video also suggested a different treatment of the visual parameters. The video's progression was defined by the protagonist's facial expression, camera distance, visual effects, as well as the use of cling wrap on the protagonist's face. The editing dynamics for this video were also derived from a manually edited video sequence.¹³⁰

Inspired by John Cage's *Music of Changes*, as well as Walter Murch's suggestions towards a notational system for film (Ondaatje 2002), the parameter models described below used a structure grounded in the principles of the *I Ching*.

The following parameter model was used in the development of the video for *Look... The Sun is Rising*. It consisted of four groups of parameters, with each parameter group containing three parameters, and each parameter had an associated binary (yes-or-no) state. For this video, the parameters and their groups were defined as follows:

Context:

- Flare (lighting);
- Extreme close-up (field size);
- Cling wrap.

¹³⁰ chuls. "The Flaming Lips - Look... The Sun is Rising", *Genero.tv*, 11 May 2013, <http://genero.tv/watch-video/36029>.

Subject:

- Eyes closed;
- Eyes open;
- Neutral (facial expression).

Effects 1:

- Strobe;
- Time displacement;
- Glitch artefact.

Effects 2:

- Composited (two or more shots are combined in a single piece of footage);
- TV static;
- Datamoshed.

With this model, each parameter could be described in the same (binary) way; only their labels were changed. This approach, however, drastically changed the way in which these parameters were considered. As an illustration, let us consider an example using field size. Rather than creating a taxonomy of various camera distances, the occurrence of the extreme close-up shot is defined as a key characteristic. The algorithm therefore decides if the next shot should be an extreme close-up or not. The sum of these decisions across multiple parameters creates a filter that can be applied to the clip library for making shot selections. Figure 7.2.27 shows an example of this type of structure in three segments to illustrate the possible progression of a parameter's dynamics throughout a video sequence.

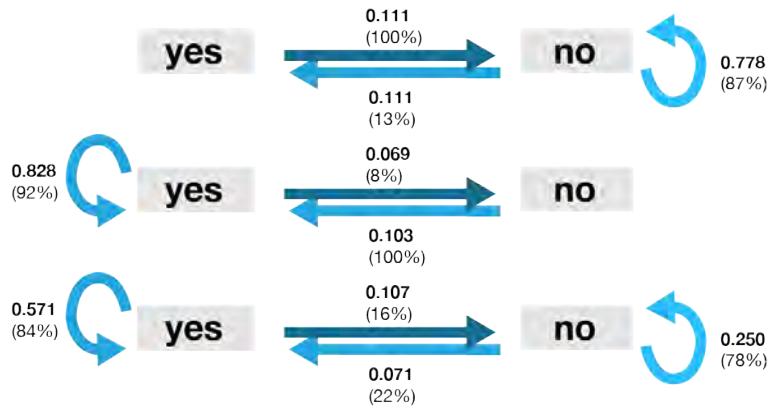


Figure 7.2.27: Example of the progression of a parameter's dynamics described using the segmented binary decision model.

Following the application of this model in *Look... The Sun is Rising*, the same process was applied to the development of the video for *Majestic*, resulting in the definition of two groups of parameters: subject and context.

Subject:

- Band;
- Kate;
- Paper silhouettes.

Context:

- Setting;
- Performance;
- Setup.

The subject group describes who is seen in a shot. For this purpose, the band was treated as a single protagonist, rather than differentiating between the appearance of individual band members, thus indicating if any band members are seen at all. The paper silhouettes were treated as a single protagonist as well. The context group in turn describes the action, or lack thereof, in a shot.

7.2.3.5 Decision Constraints

The introduction of parameter groups and binary state decisions allowed for an increased computing efficiency and coherence in the algorithmic decision process. The introduction of the binary parameter model also simplified the process of defining individual parameters. However, due to their low information content, additional conceptual input was required to define their relation to other parameters. This led to the grouping of parameters according to their logical or narrative dependencies.

Ideally, the utilisation of transition probabilities would allow for an optimal decision process, which ensures aesthetically safe choices. However, given the algorithm's variability, this was not always the case. Therefore, a function defining and mapping parameter dependencies had to be implemented to secure coherence if conflicts or dead ends arise. An example might help to illustrate this procedure.

Example

Given independent probability in a group that contains three binary parameters, we obtain 2^3 or eight possible combinations, as shown in Table 1.

Band	1	0	1	0	0	0	1	1
Kate	1	0	0	1	0	1	0	1
Silhouettes	1	0	0	0	1	1	1	0

Table 7.2.1: Possible parameter combinations for a set of three parameters.

As described in the previous section, the subject group comprises the parameters band, Kate, and paper silhouettes. Due to the availability of footage and a defined narrative progression, the editing dynamics for *Majestic* reduces the sample space of possible events from eight to six: Kate can only appear individually or with both, band and paper silhouettes, but never with only one of them. Figure 7.2.28 illustrates the parameter relationships.

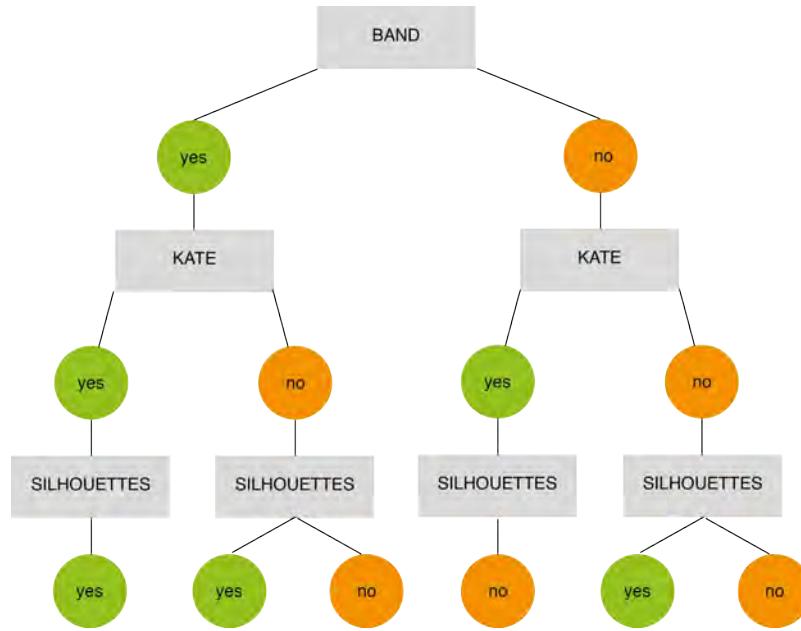


Figure 7.2.28: Subject parameter relationships in the *Majestic* music video.

In contrast, the parameters in the context group (cp. Figure 7.2.29) are all mutually exclusive, which means that only one of the parameters in the group can be true. This group is defined to be second in hierarchy, and therefore depends on the subject group. Subsequently, context events can only occur for particular combinations of events in the subject group. For example, the occurrence of the band implies that either the performance or the setting parameter is true.

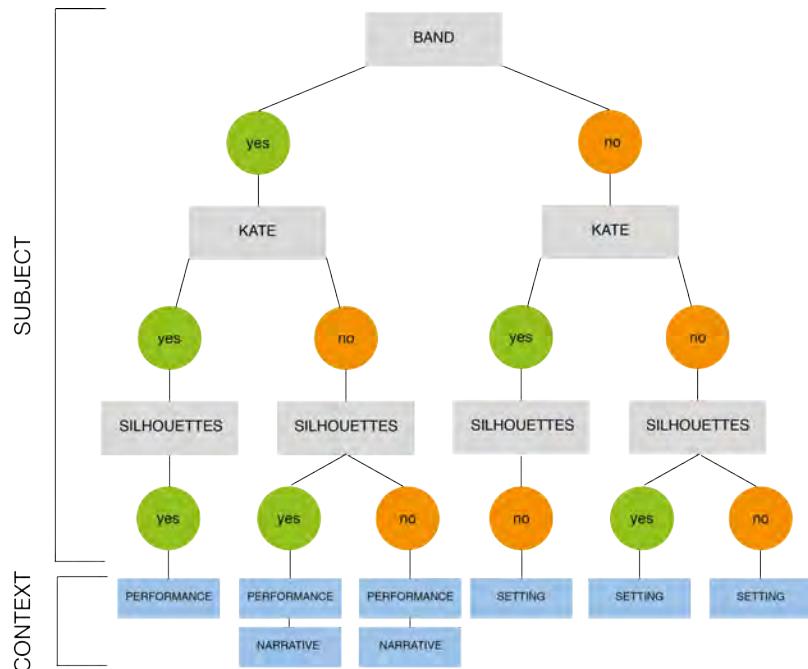


Figure 7.2.29: Subject and context parameter relationships in the *Majestic* music video.

7.2.3.6 Cutting

As opposed to *version one*, the experiments carried out during *version two* algorithmically cut video clips of varying length into appropriate shots to be placed in the sequence. This was done to provide video clips of varying durations, thereby handing over the responsibility for cutting to the algorithm to test its potential.

The first steps in the generative editing process therefore included the preparation of footage. For this purpose, all footage was closely examined and all sections consisting of appropriate material to be shown in the sequence were identified and listed in a database. This was followed by the application of colour correction, grading and visual effects.

Because *version one* used pre-cut shots, the shot duration was used as a parameter for selection alongside other visual characteristics. In *version two*, the shot duration was treated separately because the binary decision model required a different logic for this parameter. Therefore, an entirely new function was required to manage the cutting of footage.

Three time categories linked to a probabilistic structure derived from the analysis of the manually edited sequences were defined. Once the algorithmic decision-process had chosen a shot, a cutting function would first determine a duration category for this shot, and then choose a random duration from the possible time range. For *Majestic*, for instance, the following time intervals were used:

Duration 1: 0.2 – 2.0 seconds

Duration 2: 2.0 – 4.0 seconds

Duration 3: 4.0 – 9.0 seconds

The function had to check next if the chosen clip could accommodate for this new duration. If the clip was shorter than the chosen duration, or when sections of the clip had already been shown and the remaining duration was shorter than the chosen duration, then only the possible section of this shot would be shown and its duration would be used as the basis for the next cutting decision. An implementation of this function was therefore formulated as shown in Figure 7.2.30.

```

Call cut function
    Get previous time category
    Get probability values corresponding to this category
    Make parameter decision to receive new category
    If this clip shows performance
        Determine the clip's remaining duration according to the song position
    If this clip is shorter than the duration category
        Replace category according to shot duration
    Define random duration from the time range defined by the duration category

```

Figure 7.2.30: Cutting algorithm for *Dividation version two*.

7.2.3.7 Key Shots

So far, probabilistic parameter models, as well as the definition of temporal segmentation, were used to determine the logic or narrative progression of a sequence. Next, the concept of key shots was introduced in order to provide another means of control for the editor. With regard to generative music composition, the idea of key shots can be compared to Brown, Gifford and Davidson's (2015, 14) implementation of "structural tones" in their so-called "directed random walks", which help to generate more sensible musical structures and clearer musical contours.

Key shots serve to play a chosen shot at a particular time in order to (1) ensure that memorable moments in a song are always highlighted with an appropriate visual representation, and (2) to reset editing parameters according to the defined shot, which will influence the following sequence progression. Example 1.1 in Chapter 7.1.2, for instance, has shown that sequences can deviate substantially from the intended concept. This may be desirable in some cases, but in others it may lead to outcomes becoming too similar and developing a high probability of completely omitting certain types of shots. In cases where a sequence deviates from the intended concept in such a way, a key shot may serve to interrupt and redirect the generative decision process by introducing new parameter states, and therefore a new scene.

7.2.3.8 Discussion

The experiments in *version two* using *Majestic* and *Look... The Sun is Rising* built on the methods developed during the first experiments and expanded on the concepts of defining both visual parameters and the corresponding decision models.

Before defining how to edit a generative sequence from scratch, it was necessary to define methods of audio-visual synchronisation. Using a MySQL database, each piece of footage that required synchronisation was recorded with a time code, which would link the clip to a particular moment in the song. Additional functionality to query the availability of synchronised footage was therefore developed and tested in the initial experiments for *version two*.

The experiments have shown that visual parameters for music videos can vary greatly in meaning, context, and quantity. In order to develop a standardised design for a generative editing system, it was therefore necessary to devise a simple parameter model which may be applied to any visual characteristic and which can be applied to a small as well as a large number of parameters. This led to the definition of a binary parameter model, in which all visual characteristics were described as yes-or-no questions. This model therefore no longer describes parameters as simple self-contained taxonomies, but highlights the occurrence of individual important features instead.

Manually edited video sequences were created and analysed to derive probabilistic editing values reflecting the author's creative manual practice. This approach was inspired by the previous experiments with *Dividation version one*, where an existing music video was remixed. Given that future projects continue to use these probabilistic structures, future implementations of this technique will build on processes analysing manually edited sequences automatically, after parameter definitions have been made and footage annotated. At the time of this writing, however, these processes require further research to produce a feasible workflow for generative music video editing.

The binary decision model developed in *version two* was based on concepts of the *I Ching*, inspired by John Cage's *Music of Changes* and Walter Murch's implications towards cinematic notation. Therefore, parameters were placed in groups of three, corresponding to a trigram in

the *I Ching*. The grouping was done for two reasons; (1) to reduce the number of decision cycles of the editing algorithm, and (2) to establish parameter relationships.

The process has proven useful. The development of generative algorithms, as well as the adaptation of these algorithms to other music videos, appeared more intuitive when using identical parameter structures and their relationships. These methods extend the author's control over the generative editing process, in addition to the definition of the probabilistic structures.

The definition of parameter relationships served to impose constraints on the algorithmic decision process, a requirement described at the end of the experiments of *version one*.

Because editing parameters are defined individually, they must relate to other parameters in order to support accuracy and efficiency in the decision process. Constraint functions avoid impossible decisions and eliminate unnecessary decision cycles.

A further means of control over sequence progression was provided through the introduction of key shots. So far, the use of temporal segmentation served to change editing dynamics at given points throughout the sequence. Key shots provide a similar function. First, a key shot is placed to highlight a particular moment within the sequence which must remain the same throughout any iteration. Second, because algorithmic editing decisions are made individually for each cut and depend on the preceding shot, key shots can be placed to reset parameters at a particular time in the sequence to redirect the algorithmic decision process in case it deviates too far from the intended concept.

Additionally, according to ideas developed during *version one*, *version two* changed the timing of the editing process from being computed in real-time during playback to the planning of the entire sequence once the play button is pressed, before playback begins. Depending on the algorithm complexity, this can result in comparatively long waiting times before playback, but eliminates the risk of delays during playback due to the background calculation of editing decisions.

Example 2.3¹³¹ shows three generative videos created for *Majestic*.

¹³¹ Julia Stefan. "Sabre Siren – Majestic // Generative music video edit", YouTube video, 3:56, 29 September 2014, <https://www.youtube.com/watch?v=MFZ-IuHCdL8>.

7.2.4 Version Three

The fundamental ideas behind the development of *Dividation* increasingly focused on the definition of an efficient and intuitive model to describe editing dynamics using probabilistic transition structures, as well as the development of control methods.

Version three continued to explore these topics, specifically extending on the areas of cutting, shot duration, and increasing narrative detail.

Shot duration and algorithmic cutting were revisited in *version three*, on the one hand to explore methods to facilitate online streaming without the need for extensive server-side processing, and on the other hand to investigate cutting as a means of control over sequence aesthetics and progression for the author.

The basic requirements for this phase of the development of *Dividation* therefore concerned:

- Optimisation of algorithms for online streaming;
- Extending creative control;
- Developing methods for greater narrative detail;
- Exploring aesthetic possibilities.

Version two introduced a cutting method which autonomously altered the length of provided footage according to separate time range definitions and corresponding probabilistic structures. This method provided a high level of variability, but created several challenges for the loading and streaming of sequences. Furthermore, when applied to an entire sequence, this level of variability may not be desirable for many editors. Balázs (1970, 131), for instance, stated that “the art of cutting consists in the first place in determining the length of each shot”. When speaking of their practice and methods, other editors highlight the importance of guiding the viewer’s eye, and of creating a smooth flow over a series of shots (Oldham 1992, Ondaatje 2002).

Cutting itself may constitute the aspect of editing most subjective and difficult to formalise. Therefore, it was proposed that keeping this process largely in the hands of the editor would be preferable over previous approaches using algorithmic cutting for the entire music video sequence. Instead of creating clips of arbitrary durations, the editor was now required to cut a

larger amount of shots, each of them provided in its final duration. Algorithmic decision-making is then primarily used to select and place shots in the sequence. The algorithm only alters the duration of shots where synchronisation constraints require variations.

This phase of the development also focused on greater narrative detail to create a music video that corresponds to the Disrupted Narrative, outlined in Chapter 5.4.3. *Version two* had focused on sequence progressions corresponding to the Vague Narrative. These videos could be described according to the appearance and disappearance of comparatively few visual features. *Version three*, in turn, would emphasise a sequence of discrete actions, requiring a different kind of structural definitions.

The development and testing of *Dividation version three* was based on practical work, i.e., the creation of a music video for the Brisbane band Rohan and their song *Eloise*. The details of the changes to cutting and narrative requirements for this production are described in the next section.

7.2.4.1 Preparation of Footage and Cutting

Similar to previous experiments, a linear music video was manually edited and used to derive definitions for editing dynamics.¹³² Afterwards, the development of generative editing algorithms in *version three* first had to focus on the preparation of footage, which strongly influenced the progression and aesthetics of the generated outcomes because of the different approach towards cutting adopted at this stage.

The generative music videos created during *version two* consisted of 100 to 200 individual clips of arbitrary length. In these experiments, all footage that depicted smooth movements, relevant content, and no obvious errors was provided to the algorithmic editing process. The editing algorithm would then make decisions regarding the order and type of footage to place in the sequence, and individually cut each clip into shots of durations specified by the program.

¹³² AChampionSound. “ROHAN | Eloise”, YouTube video, 3:47, 3 July 2014, <https://www.youtube.com/watch?v=D7S1NfXedkc>.

In *version three*, however, a completely different approach to editing was taken. Here, all clips provided to the editing algorithm had already been manually cut into their potentially final durations. On the one hand, this lead to a complete elimination of the algorithmic cutting function, and on the other hand to a much larger number of clips provided to the program. Whereas *Majestic*, for instance, worked with a pool of approximately 200 clips, the pool for *Eloise* exceeded 800 clips. The total duration of all clips added up to almost forty minutes for *Majestic*, and about twenty minutes for *Eloise*.

Because of the increasing amount of effort required to create these shots, the generative video's variability may decrease. The examples in *version three* will illustrate that the number of clips drastically increases when using this method, but the total amount of footage provided for *Eloise* is only around half the total footage provided for *Majestic* because the individual shot durations are very short. The longest clips for *Majestic* were several minutes long, whereas the longest shots for *Eloise* were under twenty seconds.

With regard to streaming efficiency, this method proved successful. All video footage was encoded with the VP8 video codec, with the resolution set to 720 x 404 pixels and a rate of 25 frames/second. Whereas in the previous version the *Majestic* clips required a data transfer rate of 30 MB/minute or higher, resulting in frequent interruptions during clip loading unless a high-speed Internet connection was provided, in contrast, in *version three*, the *Eloise* clips required only up to 13 MB/minute and therefore the probability of playback interruptions decreased.

By removing the responsibility for cutting from the algorithm, the author regains some control handed over to the program in previous experiments. With *version three*, the author decides where each shot begins and where it ends, whereas previously achieving an appropriate cut was largely left to chance.

In order to produce a large amount of shots that correspond to the footage in the sequence, a systematic approach was selected. First, each shot of the manually edited sequence was provided. Then, for each shot, two similar shots with the same duration were created. Finally, a number of two-second long shots were cut from footage not used in the manually edited sequence, as well as several two-second long performance shots, in order to close any gaps. So far, not enough synchronised performance material had been available for the algorithm to choose from.

The duration of individual shots was only changed when the placement of performance shots required adaptation in order to achieve accurate audio-visual synchronisation.

7.2.4.2 Narrative requirements

The music video for *Eloise* was shot in two days. On the first day, the first part of the footage was recorded in the apartment shown in the music video. The second day focused on capturing close-ups of the band in a separate rehearsal space. This video was planned to feature a disrupted narrative, and it therefore displayed more narrative detail than the videos created in previous experiments. The lyrics of the song *Eloise* directly address a friend going through a bad phase, and the song tries to empathise with this person's experience. Therefore, the video was shot from a third- as well as a first-person perspective using an additional GoPro camera attached to the protagonist (Figure 7.2.31).



Figure 7.2.31: Adjustment of the GoPro camera on set.

The video depicts an entire day in the life of Eloise, the protagonist, which she spends in her home. In addition to telling this simple story, the music video also aimed to show a typical Brisbane apartment, and therefore equally focuses on the protagonist and her surroundings.

In order to preserve some accuracy towards the passing of Eloise's day it was necessary to find formal definitions that can describe a series of events, and therefore go beyond the addition and removal of several simple visual characteristics, which constituted the dominant mode of defining the sequence progression in *Majestic* and *Look... The Sun is Rising*. Similarly to *Big Jet Plane*, the narrative progression in *Eloise* is largely defined by the protagonist's location. Here, however, a much larger number of locations were needed. The following list illustrates the succession of locations as depicted in the manually edited sequence (Figure 7.2.32).

- 1) Bedroom: Eloise gets up;
- 2) Bathroom: Eloise looks at herself;
- 3) Kitchen: Eloise reaches for drinks, the band is shown behind her;
- 4) Living room: Eloise sits down to watch TV, the band makes contact;
- 5) Outdoors: Eloise runs outside.



Figure 7.2.32: The five locations shown in the music video for Eloise.

In addition to displaying a larger number of locations than *Big Jet Plane*, the transitions between these locations tend to be illustrated as well because we are following Eloise throughout her day. It is therefore more difficult to describe these locations as discrete segments. It was possible to revisit one of the first concepts introduced for *Majestic*, where initial experiments used phases to define sequence progression, but rather than relying on these discrete groups in which events could take place, a different method for describing an inherent order in the footage was designed.

7.2.4.3 Towards Narrative Coherence

With the narrative structure in *Eloise* requiring a large number of events and actions, previously developed processes—primarily based on Markov chains and probabilistic editing structures—had to be extended in *version three*. Because of its narrative content, the music video for *Eloise* requires more multi-dimensional visual parameters, which cannot be easily described as characteristics appearing and disappearing from a scene. The relatively large number of locations, respective actions, and the inherent order of these events asked for a different approach in addition to the previously defined binary parameter model.

For this purpose, each shot in the database was recorded with a time code that would determine the shot's absolute or relative position within a generated sequence (cp. Figure

7.2.33). These time codes were taken directly from, or related to, the manually edited sequence, which served as a guide for both editing dynamics and shot creation.

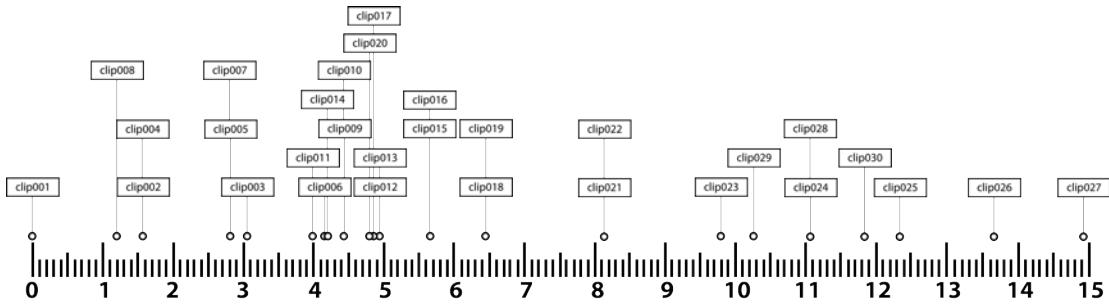


Figure 7.2.33: Illustration of the time-code based structure of the first fifteen seconds of the *Eloise* music video.

The previous experiments only made use of absolute positioning for footage, which required synchronisation, such as performance and key shots. Here, the concept of relative positioning was introduced for all other shots in order to facilitate a linear narrative arrangement corresponding to the manually edited sequence. Using the time codes recorded in the shot database, the algorithm received a list of shots ordered according to these time codes.

Because of this time ordering of shots in the database, a simpler parameter model could be used. Three parameters were defined, which indicated the occurrence of:

- Sep (the singer),
- Eloise (the protagonist), and
- the band.

Building on the concept of parameter groups introduced in *version two*, these parameters correspond to the subject group. The context group, in turn, was replaced by the method described in this chapter.

The experiments continued to use segmentation to develop editing dynamics using the binary parameter model, and to provide a further means of control for the implementation of the time-code based narrative structure. In previous experiments, the algorithm would always make a random choice from the filtered footage, which corresponded to the probabilistic editing decision. Here, an additional probability value was introduced to determine whether a decision must be made randomly or whether the linear time-based structure must be considered instead. Figure 7.2.34 describes a simple implementation of this function.

```

If this is the first shot
    Play first shot
Else
    For each parameter
        Read previous state
        Find state probabilities for next shot
        Determine next state with random number according to probabilities
        If the new state does not adhere to parameter constraints
            Transform state to match constraints
        Form a set by fetching all shots that match the new state
    Intersect three sets into final set
    If the final set is 1
        Play this shot
    Else, if the final set is larger than 1
        Determine if shot decision is random or ordered
        If the shot decision is random
            Fetch random shot from the set
        If the shot decision is ordered
            Fetch shot whose time code is closest follower of preceding shot

```

Figure 7.2.34: Shot decision algorithm using the time-code based narrative structure.

The interface developed for *version three* included options for interactivity to influence segmentation, as well as the probability of the linear narrative structure being used, and to prioritise the selection of either narrative or performance clips (Figure 7.2.35). Segmentation therefore no longer constituted a constant definition linked to musical sections, for instance, but was exclusively used in intervals of a given duration. Here, segmentation was therefore used to increase or decrease the variability of the editing dynamics. As a result, longer segments would lead to a more open decision process because more footage was available for any given segment, and shorter segments may become more temporally accurate, but could also increase the chance of situations with impossible solutions due to the lack of available footage and reduced structural possibilities. *Version three's* time-code-based narrative structure was influenced by changing the additional probability value, whereas the priority options influenced the pre-defined values of the binary parameter model.

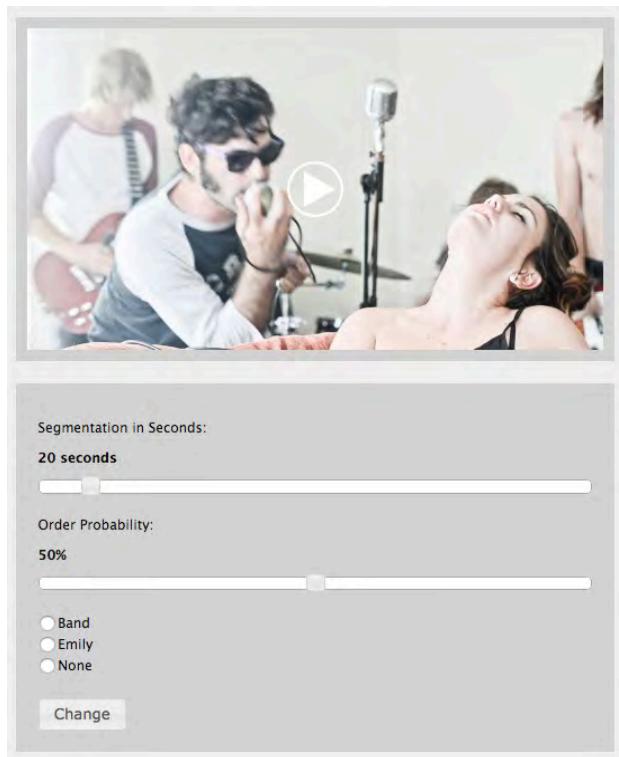


Figure 7.2.35: Interactive options for the *Eloise* music video.

Figure 7.2.36 further highlights the influence of segmentation particularly on relative positioning. Here, each clip strictly belongs to a particular segment.

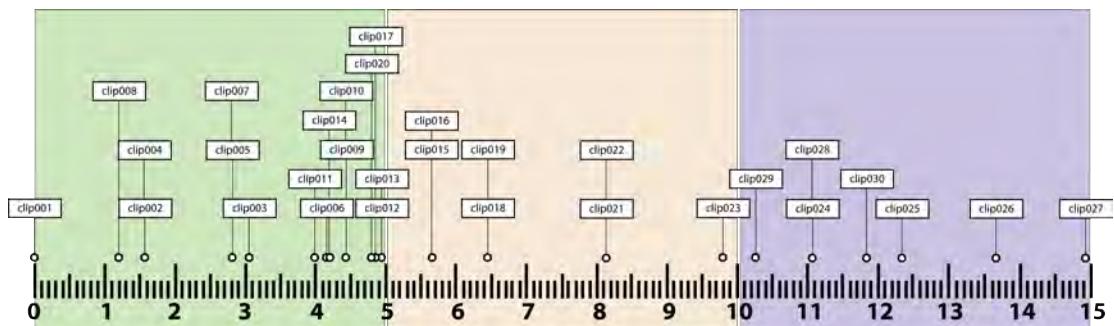


Figure 7.2.36: Illustration of the time-code based structure of the first fifteen seconds of the *Eloise* music video using five-second segments.

7.2.4.4 Discussion

Dividation version three served to further extend the methods developed in the previous stages, and to explore the generative editing of narrative structures more deeply. Additionally, online streaming was considered.

Appendix C lists the source code for version three's generative editing algorithm written in PHP. Example 3.1¹³³ shows three generated videos for version three and a live example of the generative editing algorithm in action can be found in Example 3.2¹³⁴.

First, the generative editing process was transformed by removing the algorithmic cutting function, and by changing the way video footage was prepared prior to the editing. For this purpose, the editor was now required to cut the footage into finalised shots, which may be shown in their entirety, unless the need for better synchronisation required temporal adjustments.

From a technical point of view, this method was successful in reducing the amount of footage loaded into the browser during playback, without requiring additional server-side processing of video footage. It reduced the amount of transferable data to less than half the amount required in *version two*.

An obvious drawback of the method, however, is the increase in workload required to prepare the footage. Because each provided shot was likely to be significantly shorter than clips provided using the previous method, a much larger number of individual shots was needed to provide enough footage for generative editing. However, the method also allowed for a greater control by the editor, who can now specify shot durations, rather than having to rely on the algorithm to make the right cut. This is an important consideration because determining the precise start and end of each shot in a sequence is a vital aspect of editing.

At the time of this writing, all of the footage has been prepared manually. Therefore, a requirement for further research is the creation of appropriate interfaces and automated processes which support the editor's creative practice regarding these editing decisions, while allowing the editor to retain much of his or her control over the content provided as the basis for the generated outcomes.

A new method to extend narrative complexity was introduced in addition to the probabilistic binary parameter model developed in the previous version. This method was based on defining a timeline by annotating each shot with a start time. Previously, only footage which

¹³³ Julia Stefan. "Dividation – Version 3, Eloise", YouTube video, 3:38, 1 March 2015, <https://www.youtube.com/watch?v=uQYpIkePfTE>.

¹³⁴ Julia Stefan. "Eloise". <http://juliastefan.com/eloise/>.

required synchronisation, e.g., musical performance and key shots were treated in this way. *Version three* differentiated between absolute and relative positioning. Synchronised shots therefore required absolute positioning; i.e, they cannot be played at any other time. Temporal positioning of all other footage was considered to be relative because the time codes provided for these shots only determined the segment in which they were available for playback, as well as their original order in relation to other shots. Compared to previous versions, with this temporal information it was possible to algorithmically define a narrative structure while reducing the number of visual parameters required.

In the software, the implementation of this time-code based narrative structure is linked to a probability value, which determines the degree of adherence to a clip's timeline location in the algorithmic decision process. Thereby the editor is able to define the structure's priority for the use of defined clip positions over the use of random values.

The reason for introducing this method was the requirement for developing intuitive processes supporting the author's creative practice, rather than prescribing methods which may yield accurate sequences, but would likely limit the author's creativity. The experiments undertaken with *Dividation* sought to design a workflow and methods orientated towards the potential needs of the editor according to my own experience, as well as descriptions of the practice by various editors (Oldham 1992, Oldham 2012), and fulfil creative requirements for sequence creation. Future research will work more closely with editors to test usability and requirements in further practice.

The final generated outcomes for *Eloise* were in good agreement with the original concept depending on the chosen settings. In addition, different approaches in the development of *Dividation* have also revealed changes in editing aesthetics. The manually edited music video sequence for *Eloise* was partly based on the use of repetition and jump cuts. This stylistic choice was reflected in the generative editing process by providing a large number of similar shots, which can be used to display overlapping, repetitive action. By adopting a time-code-based narrative structure, it is therefore possible to recreate this editing style in a relatively controlled fashion because series of shots can be played according to their previously defined temporal order.

Version three therefore also highlights the need for different generative editing methods for different purposes. While the cutting and segment boundary methods introduced here may be

more feasible for the editing of large parts of the sequence, the cutting function developed in *version two* may be applied selectively during particular parts of a music video sequence, for instance, the stroboscopic inter-cutting between two shots shown in the music video for *Gangnam Style*, described in Chapter 5.2.1. The individual shots used for this technique can be as short as two frames. For the generative editing process, it may not be feasible to provide shots of such a short duration, but by using the cutting method introduced for *version two*, these effects may be recreated by algorithmically cutting between shots.

Future research on generative music video editing methods will have to test a larger number of different narrative structures and examine the influence of different methods and their combination on generated sequence outcomes.

7.2.5 Author Survey

The second survey conducted for this research targeted video authors, e.g., directors and editors (cp. Appendix D). Here, a more qualitative approach was taken by focusing on fewer participants and asking for detailed commentary instead of an evaluation using rating scales. The survey was again provided online and participants remained anonymous.

The survey sought to clarify how creatives in the field view generated results in order to (1) evaluate the current state of the generative editing algorithms developed for *Dividation*, and (2) to learn more about their application in practice.

Participants for this study were approached according to their professional background and experience in the field of video editing or video production in general. They were therefore required to have at least one year of experience in the field in order to qualify for the survey.

Furthermore, they were informed regarding the general process of generative editing applied in the music video for *Eloise* (cp. Chapter 7.2.4), and were then asked to play and watch the video three times, effectively resulting in three unique sequences provided to each participant. Then, the participants were asked to answer the following questions:

1. What were your thoughts of the different versions of this video?
2. Please describe your impressions of the video editing.
3. Were there sections in any of the videos that you liked particularly, with regard to the editing? If yes, please explain.
4. How do you think the editing could be improved?
5. How do you think the editing would affect a music video audience?

Seven participants answered the survey and qualified for analysis. In the following sections the participants' context and their perceptions of the videos will be briefly discussed. Afterwards, a summary of the results will be given, which will first highlight the music video's editing style, the influence of different aspects of visual content on the perceived quality of the music video, and the role of variation created by the generative editing algorithm. Finally, some accomplishments of the generative video processes and areas requiring further improvement will be discussed.

7.2.5.1 Participants

Participant	Gender	Age	Experience	Comments
1	Male	31-40	4-8	The participant indicated that the video was consistent with current music video trends.
2	Male	41-50	16-30	This participant expressed a strong dislike for the musical as well as visual content of the video. He indicated that the editing is typical for music video with a strong focus on rhythm but a lack of meaning.
3	Male	31-40	1-3	The participant expressed an interest in the nature of the experiment, but pointed out that the video lacked meaning. He mentioned several areas of improvement for the editing, but also suggested that these may not matter to an audience.
4	Male	41-50	9-15	The participant stated that he enjoyed watching variations of the video, and that the type of editing is suitable for this video and type of narrative.
5	Female	21-25	1-3	The participant liked the video from both a narrative and a stylistic point of view. She did not observe strong differences between the three

				iterations and identified issues in synchronisation and continuity.
6	Female	41-50	9-15	The participant did not like the video but indicated that some iterations told the story better than others. She described the videos as repetitive and pointed out issues in continuity and synchronisation.
7	Male	31-40	4-8	The participant stated that the three iterations showed differences in narrative focus and indicated that he therefore preferred some iterations over others. Synchronisation was again pointed out as an issue.

Table 7.2.2: Participant overview.

7.2.5.2 *Editing Style*

The participants described the editing as “jump cutting”, “staccato”, “erratic” and “repetitive”. Some mentioned that this might be an appropriate style for this type of song and video, but it was also pointed out that this characteristic may not work for other formats and types of narrative. Some participants enjoyed this style of editing, some did not enjoy it. One participant mentioned that random-seeming cuts disrupted his attention while viewing the video, which might lead to the impression that the editing overall is not refined or that “someone” with little editing experience has cut the sequence. A second participants pointed out that he particularly liked the “time-shifting” effect created by the editing algorithm. Another participant described the use of jump cuts to be an effective method to make the video “edgy”. This participant also mentioned that she enjoyed the crosscutting between Eloise’s story and the close-ups of the band’s performance. The repetitive editing of Eloise’s story has been referred to as both positive and negative. Positive views regard this technique as effective in pointing out the protagonist’s repetitive and dull activities as part of the narrative.

Several participants also stated preferences for the editing of certain sections of the generated music videos, especially concerning a section of the music video where the editing will usually increase in pace as the music intensifies, cutting back and forth between the protagonist and the musicians.

7.2.5.3 Influence of Narrative and Audio-visual Content on the Perception of Quality

Descriptions of the editing were more negative when participants disliked the content, e.g., the song itself, as well as the visual footage and the overall narrative. Concerning the visuals and the narrative, some participants perceived the video to be “meaningless”. These participants stated (1) that the footage and the narrative should be more complex and expressive and (2) that every cut should be a meaningful juxtaposition between shots. In addition, it was also stated that some shots should have been chosen more carefully, especially when the clip showed more than “the best part of the shot”.

Some critical reception can be correlated to an aversion of popular music videos in general. One participant stated that he did not enjoy the video, but that the style of editing shown is common for many music videos. Although he perceived the video to lack meaning he said, “it worked rhythmically”.

In contrast, some participants also stated that all of the generated sequences displayed “consistent narrative arcs” and that, for videos without a complex narrative, this method “would be perfect”.

7.2.5.4 Variation

The perception of variation between the videos caused by repeated generation of the video sequences was mixed. One participant enjoyed the “subtle differences” with each new generation. Some participants stated that the sequences were very similar. In this context, it was said that “there seemed to be little difference between the three cuts, style wise” and that “the storyline wasn’t effected by through out [sic] the different versions”. Other participants perceived a focus on different narrative aspects in each video iteration, for instance, stating that it was “fascinating to see 3 subtly different takes on the same topic”. When discussing the different versions, one participant pointed out that “the second seemed to tell the story the best”. Another participant elaborated: “Overall, the pacing seemed similar between them, although the first was possibly the slower version. Each had a different mix of priorities with the band vs. the girl, as well as the lead singer vs. the rest of the band”.

7.2.5.5 Discussion

The overall responses to these generative music videos ranged from seeing them as random, disengaging, and unlikeable to the perception of consistent narrative arcs and effectively edited, edgy sequences.

Similarly to the previous survey, several parameters render the evaluation of the algorithm's quality difficult. First, some viewers dislike the music video format in general, or the visual content of the music video in particular. The interpretation of these survey results therefore needs to consider the aesthetic preferences of authors and their priorities regarding what constitutes effective editing and the viewing of music video sequences.

Several participants mentioned that the type of editing applied in the video was appropriate for this style of music, consistent with current music video trends and that it creates rhythm, although it may be lacking in meaning. Although they observed issues in the editing some participants also mentioned that this might neither be noticed by audiences nor cause them to disengage.

Some comments were made concerning variation in this demonstration. One participant asserted that he enjoyed the changing sequences; another said that he could see clear narrative differences, and another participant mentioned that these changes were (perhaps too) subtle. The latter shows that the applied editing method was effective in recreating the defined concept for the video's narrative. Although an interactive interface had been created for *version three* of the software that generated these music videos, it was not used in the survey because the priority was to gather information on the effect of the algorithmic outcomes at fixed, pre-defined settings. A general lack of variation perceived in the sequence was therefore not surprising.

The participants' feedback confirmed several areas requiring further improvement for future versions of the editing algorithms. It was mentioned that there are obvious "bad cuts" from time to time. A bad cut, in this context, can be considered as a shot that either displays not enough information, or too much. Walter Murch, for instance, described the ideal length of a shot as the moment "it has revealed everything that it's going to reveal, in its fullness, without being overripe. If you end the shot too soon, you have the equivalent of youth cut off in its bloom. Its potential is unrealised. If you hold a shot too long, things tend to putrefy"

(Ondaatje 2002, 267). Brooks (2010, 796) confirmed that editing is perceived as bad, when its timing is “somehow ‘off’”.

Planning and preparing sequences in advance may largely solve this issue. The design of interfaces and workflows for *Dividation* will require a process that makes the creation of footage more efficient in order to increase its quality. *Dividation version three* required to manually cut around 800 shots. This is not a trivial task, and without an appropriate tool to create all of these shots, it is likely that either the time required to create these shots will be unreasonably long, or the quality of individual shots may be insufficient. In order to introduce the method in *version three*, a balance between these two aspects was sought. While some inappropriate shots were filtered out during development, some remained in the shot pool for experimental purposes. Future research will necessarily have to include finding processes that can facilitate the task of shot creation.

Given the role of randomness in the automatic generation of these sequences, there will probably always be a risk of an inappropriate juxtaposition of shots. Further research will also require a continuous evaluation of current generative editing methods, and finding new means of control over sequence outcomes.

In some parts of a sequence, the editing may not always reflect the pace of the music. In the generative music video created for *Eloise*, this tends to happen in a particular section of the video when the music intensifies, with the editing meant to crosscut between the protagonist and the musicians in fast succession. A participant positively mentioned this particular technique when it was achieved. But if the algorithm deviates from this concept, participants may perceive the timing of the editing to be wrong. A particular challenge for this technique lies in the short duration of the shots needed to achieve the desired effect. A large number of short shots, often only comprising several frames, must therefore be provided. This is further complicated by the requirement for synchronised shots, which can only be played at certain times. For the purpose of experimentation, some definitions were deliberately left open in order to allow for variation of these techniques and an evaluation of their perceived quality. To guarantee that this type of effect can be achieved, several methods may be applied. First, the chance for the effect to occur can be increased by providing a larger amount of appropriate footage for this section of the music video. Second, it is possible to implement stricter temporal definitions, which, similarly to the concept of key shots, cannot be changed or avoided. Third, future experiments may be created with the introduction of a new

technique utilising the cutting method introduced in Chapter 7.2.3 for specific sections of a music video only.

Finally, a big concern in these experiments is audio-visual synchronisation, which in many cases is correlated to download speed. While *Dividation version three* focused on improving streaming capabilities through the algorithm design itself, the technical implementation of generative music videos online is a complex process and one that needs to be investigated in its own right in future research and development.

8 Conclusion

"There are no muses. All the arts were invented by humans. However, in many cases – such as music and dance – that invention [...] has been forgotten, lost, as it were, in history. Yet in other cases, arts have been self-consciously created. Sometimes this has been the result of hybridization – the combining of pre-existing artforms, as in opera. Or, as in the case of film, video and photography, artforms have been erected upon the technological discovery of new media."

Carroll (1996, 3)

Inspired by evolving technologies and current cultural developments, this thesis provided an overview of an emerging new media form and creative practice. Generative music video editing constitutes a process that can harness these emerging technologies and provide a form that corresponds to these changes in the media landscape.

In order to establish music video's aesthetic requirements, this thesis educed principles for the editing of this form from (1) literature, (2) a comprehensive analysis that served to compare a variety of music videos and derive conclusive definitions informing the priorities of the following system development, and (3) the author's own creative practice.

The analysis confirmed that music video is a highly heterogeneous video form, and that these videos cannot be described by a potential set of templates. Some common editing patterns and techniques were identified for linear music videos, but interactive music videos so far make use of varying methods for the creation and presentation of their visuals, which cannot necessarily be categorised.

Dividation, a prototype system for generative music video editing, was introduced as a novel approach to define generative editing as a creative practice based on the requirements of both author and system in order to generate adequate results.

The development of *Dividation* aimed to provide a process for non-linear video creation that could generate a high level of variability without the requirement for user input. For this purpose, *Dividation* has been designed as a tool for professional editors to support their creative practice.

The research that informed the development of *Dividation* has therefore begun to synthesise music video editing trends, and eventually provided a process that generates sequences utilising probabilistic methods, e.g., Markov chains, for varying the temporal structuring of video clips.

The thesis has stressed the importance of such a tool and has investigated both its technical and cultural implementation.

8.1 The Evolution of Audio-visual Media

Our experience with audio-visual forms has predominantly been of a linear nature. So far, film or video sequences have been fixed structures, assembled by an editor before presenting them to an audience. This project was motivated by the question whether this mode of operation is inherent to the way we create and consume audio-visual sequences, or whether it is a convention born out of the possibilities available to us?

Invented in the late nineteenth century, film is considered to be a very young art form. Its technological history has seen an evolution from black and white, and often non-narrative silent films, commonly exhibited at vaudeville exhibitions, to colourful high definition digital formats offering surround sound for cinematic exhibition or streaming at home, among other possibilities for viewing. From a mainstream perspective, the viewing of these films and television programs is considered a passive pastime in which a narrative unfolds. If, or to what extent, viewers may want to influence the progression of these sequences remains questionable. However, moving image forms in online media have begun to successfully make use of interactivity and non-linear sequencing. This research project aimed to investigate further potentials of this trend.

Music video, officially established in the 1980s, has become a popular format in online media. In the past years, its subcategory of interactive music videos has been a popular format for technological experimentation.

Forms like the narrative feature film and television drama may remain part of popular culture without seeing major transformation regarding the functionality of narrative progression. Online media, however, have introduced an inherently different type of media consumption.

It is reasonable to assume that new formats that correspond to the requirements of changing consumption will exist in parallel to other forms, depending on the media platform.

8.2 The Development of Dividation

The first chapters in this thesis covered three essential questions regarding the research on generative music video editing which can be summarised as the why, how, and what.

Why was the research on generative editing methods conducted? By discussing the development of media technologies, characteristics of online media, as well as the rising importance of audio-visual formats in these media, Chapter 2 introduced the dominant reasons for the development of generative editing methods. Online media objects are becoming increasingly multi-dimensional, offering users more possibilities to explore varying aspects of a text, thereby maintaining engagement and increasing awareness.

To date, in online music videos, this multi-dimensionality has largely been realised through the implementation of interactivity, which allows the viewer to influence sequence progression and visual appearance. In addition to providing audiences with a new media text to engage with, with these methods designers and filmmakers have had new possibilities to enhance and extend their own work.

In this research, automation was highlighted as an affordance of evolving technologies and as a method to achieve increasing variability in audio-visual texts. Rather than constituting a method of providing efficiency, automation was therefore described as a means to free the creative mind from repeating processes that in some cases may better be managed by computational systems.

Investigations into film and video editing, however, identified editing as a process that cannot be rationalised or summarised in a notational system because of its comparative young age as an established technology and practice, as opposed to many musical forms, for instance. Music video editing, therefore, remains an intuitive and comparatively unformulated practice.

How, then, could generative editing function anyway? In order to find possible answers to this question, the research turned towards art, particularly focusing on areas like generative art and metacreation. These art movements highlight the role and varying levels of artificial

intelligence, randomness, and—therefore—technological autonomy. They further underscore the importance of the relationship between the artist and his or her tools for the creation of artworks and media artefacts. While this thesis has not endeavoured to provide answers to the question of what is, and what is not art, it has discussed the role of randomness and systematic procedures in relation to creativity and human expression. Especially since the twenty-first century, artists have begun to utilise computational variability to bring the work closer to a natural dynamic outside the area of human influence (Candy and Edmonds 2002). The artist therefore provides a conceptual framework for the work, but may, to some degree, remove him- or herself from many of the final decision processes that create the final outcome.

Finally, the development of *Dividation* built upon studies on music video itself and helped to investigate the question: What are the aesthetic definitions that need to be considered for the design of generative editing methods? For this purpose, the production of music video was examined in two chapters, first by exploring existing case studies and definitions of music video, and second by creating a new set of case studies, specifically aimed at interrogating computer-assisted editing practices.

The music video analysis provided a means to address different types of music video by focusing on their progression and narrative content. This informed the development of *Dividation* by serving to differentiate between varying types of music videos and to identify essential requirements in the design of this system. A highly narrative video, therefore, requires a different functionality compared to a video stronger relying on visual rhythm and abstract concepts with regard to its progression.

The theoretical research therefore suggested the development of *Dividation* to focus on providing an open system, in which editors are free to define a music video's progression according to the individual concepts that apply to any given music video, rather than relying on particular categories or a combination of pre-defined techniques.

8.3 The Definition of a New Creative Practice

Within this matrix of music video's cultural, artistic, and medium-specific background, I developed *Dividation* by growing and creatively evolving with it. I taught the system how to generate sequences, and it taught me how this automated process may be used creatively.

Designing generative methods for creative practice can be a risky endeavour because it is difficult to not always treat them as standing in the shadows of those practices from which they are derived. Sometimes, generative methods for media creation may not appear to be very effective when compared to their traditional origins.

Generative music video editing with *Dividation* is therefore based on traditional editing. At this stage of its development, it mainly tries to imitate the resulting aesthetics of this practice in order to elicit a sense of coherence when viewing these sequences, but the methods are not (yet) designed to replace or improve the traditional practice as such. The development of *Dividation* has aimed to provide a new practice that makes use of available and emerging technologies by adding the element of variability, and therefore, by extension, chance and automation to this practice, in order to provide a new experience for viewers. Necessarily, these aspects significantly transform the editor's creative practice.

First, the production of a generative video will need to accommodate for its planned variability. The planning phase has to include concepts for desired outcomes. For *Dividation*, so far, lists of relevant shots and scenes were assembled before production, and some definitions regarding sequence progression were determined. Still, much of the work was done in an exploratory fashion. Further research will be required to investigate more closely the possible planning processes that support the practice of generative editing.

Once the footage has been produced, its treatment must constitute an in-depth preparation process, which consists of carefully reviewing, cutting, analysing, and annotating the footage. This preparation phase is vital. It defines the content of individual shots, their structuring, and therefore the appearance of the entire sequence, which largely depends on identifying visual characteristics driving the progression of a sequence.

In identifying these visual characteristics, generative editing may become a more rational process than traditional editing because editors must be able to clearly articulate which visual features determine a video's logical progression. Music video, in this context, may be based on different narrative concepts. But a large number of these videos will depend on the repetition of visual elements, or patterns somehow related to the musical structure, and the creation of audio-visual rhythm. The basic generative editing methods were orientated towards this characteristic. *Dividation*, therefore, first and foremost provided the means for the creation of

audio-visual rhythm, which was then supported by other methods based on temporal definitions to influence how these sequences produce meaning.

Most of the values used for determining the sequence structures in the algorithm were derived from manually edited sequences. During the early stages of the development of *Dividation* a method was tested in which the editor would simply annotate footage in groups and phases. However, it soon became apparent that the process of analysing an editor's practice yielded more information and values to guide the design of the generative process.

While the development of *Dividation* and the associated research revealed many areas requiring further research and improvement, the prototype system discussed in this thesis has successfully managed to create coherent music video sequences that adhere to pre-defined structures for audio-visual rhythm and narrative logic.

In addition, the development of *Dividation* has opened up the potential of establishing a new creative practice of generative music video editing, which may further inform our understanding of cinematic or videographic processes and notation.

This, however, does not mean that generative film and video editing should necessarily be regarded as a rule-governed and potentially tightly controlled process. Instead, it needs to be introduced as a means to incorporate new ideas into this creative practice according to our current and currently evolving technologies. The role of control in generative art is essential. It must be maintained, transformed, and sometimes even transferred to further allow us to explore our creative potential with the help of our tools.

Apart from technical considerations regarding the realisation of these tools from a design perspective, the greatest challenge in the creation of this process may be its establishment beyond serving the function of mere reproduction of sequences we can already create quite well with manual editing. This cannot be the sole purpose of a system like *Dividation* because the direct human influence, as well as the subjective judgment on what makes a particular cut good or effective, is likely the single most vital aspect of the practice.

What *Dividation* can do and has done, however, is to explore in what other ways we can use audio-visual material creatively, and how the outcomes of such new practices might affect and influence audiences.

This research is therefore a starting point for the future definition of the practice of generative editing and will allow further investigations of this new practice based on the methods provided in this thesis.

8.4 Future Research

Previous sections of this conclusion have mentioned several opportunities for future research. This section will elaborate these opportunities and introduce some more. During the development of generative editing methods with *Dividation*, so far, the focus was on the design of guiding principles that define the nature of the practice itself. Future research could also focus on technical optimisation with regard to processing speed and accuracy, online streaming, and both platform and software design.

It is likely that future research will continue to build generative sequences on values derived from manually edited sequences, and that the design of authoring interfaces and tools will be orientated towards making this a more user-friendly process, e.g., via automated functions for the analysis and generation of probabilistic models. There is still work to be done concerning the design of authoring tools and methods for online presentation of generative videos and their parametric control by the target audience.

Future development could also explore certain methods of implementing artificial intelligence and the understandings of complex systems, including concepts such as planning, scheduling, learning, and software approaches such as multi-agent systems and evolutionary computing.

The theory of complex systems could be considered in order to better understand the complex mechanisms behind this creative practice. On the one hand, this would serve to maintain a level of creative control. On the other hand, it might allow for a better understanding of the art of film and video editing itself, which so far remains a highly intuitive practice.

Flake (1998, 231) described complex systems as “things that consist of many similar and simple parts” whose underlying behaviour “is easily understood, while the behaviour of the system as a whole defies simple explanation”. Moving further away from linguistic, or semiological, and structuralist approaches applied to cinema and film theory in order to explain the dynamics of editing (Bordwell 1985), complex system theory may further educate

our understanding of music video editing in order to improve the design of generative editing algorithms.

The study of complex systems may influence the creative processes in the practice of generative editing itself. The cultural and technological history of music, for instance, might serve as a source of insight, where complex systems research has further influenced creative practice (Burraston 2007). By highlighting the use of cellular automata, Burraston and Edmonds (2005, 165) described complex systems as producing “global behaviour based on the interactions of simple units”. While we have yet to define potential rule systems or concrete notational frameworks for music video editing, it is possible that these systems may be applied to the practice of editing in the future as well.

The use of cellular automata, among other systems, would also highlight the importance of emergent properties in the generated output. An emergent property refers to the coherent collective phenomena which may occur in complex systems (Coveney and Highfield 1995). So far, experiments on the use of emergence in creative computation have largely dealt with issues of control in a practice depending on giving up a large amount of control. This is a common challenge in generative art (McCormack and Dorin 2001) and the role, as well as the potential influence of emergence on the generative music video editing process, could form an integral part of future experiments that consider the balance between openness and rigidity of a generative editing system. A certain level of emergence could be seen as essential to the generative editing practice. And yet, the editor’s guidance and creative concepts must be maintained as well.

While largely based on the idea of defining an outcome’s ‘fitness’, evolutionary computing might be used to learn more about optimal editing patterns. However, evolutionary computing has traditionally been applied as a method of optimisation when goals are clearly defined. This characteristic has to be considered carefully for use in real-time online systems that require a fast assembly, short loading times and seamless playback. Most systems utilising evolutionary computing will produce a large number of possible outcomes, which are then filtered selectively and transformed according to automatic mutations or user choice (Chambel et al. 2007, Romero and Machado 2008). As a result, these techniques may only find application in the preparation process of a generative music video.

Another area of investigation relevant to future research in this context includes planning and intelligent agents. Smith, Frank and Jónsson (2000, 48) described planning as a “synthesis task”, which “involves formulating a course of action to achieve some desired objective or objectives”. This sounds similar to the concept of evolutionary fitness, but a different process is used.

As an example, an approach based on planning and intelligent agents was used for Nack’s (1996) automated editing software, *AUTEUR* (cp. Chapter 3.3). Nack defined a detailed model for knowledge representation including twenty-six rules that specify how a humorous scene can be constructed. Nack’s multi-planner-based architecture was designed to manage aspects of the editing process such as narrative continuity, complex action and stylistic features regarding genre and editing techniques.

Systems like *AUTEUR* depend on explicit rules for sequence creation. *Dividation*, however, will likely continue to rely on a less formulaic and more open-ended approach, which depends as much on the definition of editing dynamics as on the way footage is prepared, as well as the amount and type of footage supplied for a given music video.

Perhaps more important than the greater field of planning may be the area of scheduling, which is mainly concerned with providing an optimal order of actions or events (Smith, Frank and Jonssón 2000).

Agent theory, in comparison, usually focuses on issues of autonomy, social ability, reactivity and pro-activeness (Wooldridge and Jennings 1995). Jennings, Sycara, and Wooldridge (1998) used three key concepts for the definition of agents: An agent is situated in a particular environment, and it performs autonomous as well as flexible actions. While the present design aimed to limit computational autonomy, it is possible that future research will enable editors to define generative editing dynamics through the use of computational agents. Of particular interest in this context may be the concept of learning interface agents, introduced by Lashkari, Metral, and Maes (1998). Using similar methods, it may be possible to provide a process that can first automatically synthesise an editor’s practice and therefore learn about this practice. ‘Deep learning’ represents another concept that needs to be investigated in this context. When implemented via online platforms, agents may even be allowed to learn from each other, and therefore provide a process through which music-video-specific notation may be approached in the future.

The potential of implementing generative editing methods in a full-fledged software tool and online platform remains subject to future research. It is my hope that the foundations laid in this thesis can benefit further development of this creative practice.

References

- Abbott, H. Porter. 2008. *The Cambridge Introduction to Narrative*. Cambridge: Cambridge University Press.
- Ahanger, Gulrukh. 1999. "Techniques for Automatic Digital Video Composition." PhD diss., Boston University.
- Alper, Meryl. 2013. "War on Instagram: Framing Conflict Photojournalism with Mobile Photography Apps." *New Media & Society* 16: 1233-1248.
- Altman, Rick. 2004. *Silent Film Sound*. New York: Columbia University Press.
- Armes, Roy. 1981. *The Films of Alain Robbe-Grillet*. Amsterdam: J. Benjamins.
- Ascher, Steven and Edward Pincus. 2007. *The Filmmaker's Handbook: A Comprehensive Guide for the Digital Age*. New York: Plume.
- Ascott, Roy. 2003. "The Construction of Change." In *The New Media Reader*, edited by Noah Wardrip-Fruin and Nick Montfort, 128-145. Cambridge: MIT Press.
- Austerlitz, Saul. 2007. *Money for Nothing: A History of the Music Video from the Beatles to the White Stripes*. New York: Continuum.
- . 2010. "'Don't You Cry Tonight' – Guns N' Roses and the Music Video Folly." In *Rewind, Play, Fast Forward. The Past, Present and Future of the Music Video*, edited by Henry Keazor and Thorsten Wübbena, 35-39. Bielefeld: transcript.
- Balázs, Béla. 1970. *Theory of the Film: Character and Growth of a New Art*. New York: Dover Publications, Inc.
- Barker, Timothy. 2012. "Images and Eventfulness: Expanded Cinema and Experimental Research at the University of New South Wales." *Studies in Australasian Cinema* 6: 111-123.
- Bazin, André and Hugh Gray. 1960. "The Ontology of the Photographic Image." *Film Quarterly* 13: 4-9.
- Bednarik, Robert G. 2003. "A Figurine from the African Acheulian." *Current Anthropology* 44: 405-413.
- Ben Shaul, Nitzan. 2008. *Hyper-Narrative Interactive Cinema: Problems and Solutions*. Amsterdam: Editions Rodopi.
- Bennett, Deborah J. 1998. *Randomness*. Cambridge: Harvard University Press.
- Berger, Jonah and Katherine L. Milkman. 2012. "What Makes Online Content Viral?" *Journal of Marketing Research* 49: 192-205.
- Bizzocchi, Jim. 2011. "Re:Cycle – A Generative Ambient Video Engine", in *Entertainment Computing – ICEC 2011*, edited by Junia Coutinho Anacleto, Sidney Fels, Nicholas Graham, Bill Kapralos, Magy Saif El-Nasr and Kevin Stanley, 354-357. Berlin: Springer.
- Björnberg, Alf. 1994. "Structural Relationships of Music and Images in Music Videos." *Popular Music* 13: 51-74.
- Bocconi, Stefano and Frank Nack. 2004. "VOX POPULI: Automatic Generation of Biased Video Sequences." In *SRMC '04 Proceedings of the 1st ACM workshop on Story Representation, Mechanism and Context*: 9-16.

- Bocconi, Stefano, Frank Nack and Lynda Hardman. 2005. "Using Rhetorical Annotations for Generating Video Documentaries." In *Proceedings of the IEEE International Conference on Multimedia and Expo*, IEEE.
- Boddy, William. 2004. "Interactive Television and Advertising Form in Contemporary U.S. Television." In *Television After TV: Essays on a Medium in Transition*, edited by Lynn Spigel and Jan Olsson, 113-132. Durham: Duke University Press.
- Boden, Margaret A. and Ernest A. Edmonds. 2009. "What is Generative Art?" *Digital Creativity* 20: 21-46.
- Bogue, Ronald. 2003. *Deleuze on Cinema*. New York: Routledge.
- Bolter, Jay David and Richard Grusin. 1999 *Remediation: Understanding New Media*. Cambridge: MIT Press.
- Bordwell, David. 1972. "The Idea of Montage in Soviet Art and Film." *Cinema Journal* 11: 9-17.
- . 1985. *Narration in the Fiction Film*. Madison: University of Wisconsin Press.
- . 2005. *The Cinema of Eisenstein*. New York: Routledge.
- Bordwell, David and Kristin Thompson. 2010. *Film Art: An Introduction*. New York: McGraw-Hill.
- Bown, Oliver. 2012. "Generative and Adaptive Creativity: A Unified Approach to Creativity in Nature, Humans and Machines." In *Computers and Creativity*, edited by Jon McCormack and Mark d'Inverno, 361-381. Berlin: Springer.
- Branigan, Edward. 1992. *Narrative Comprehension and Film*. New York: Routledge.
- Brooks, Kevin Michael. 1999. "Metalinear Cinematic Narrative: Theory, Process and Tool." PhD diss., Massachusetts Institute of Technology.
- Brown, Andrew R. 1999. "Tools and Outcomes: Computer Music Systems and Musical Directions." In *Imaginary Space: The proceedings of the Australasian Computer Music Conference*, 16-22. Wellington: The Australasian Computer Music Association.
- . 2012. "Creative Partnerships with Technology: How Creativity is Enhanced Through Interactions with Generative Computational Systems." In *Proceedings of the Eighth AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment (AIIDE'12)*, edited by Philippe Pasquier, Arne Eigenfeldt and Oliver Bown, 14-20. Palo Alto: AAAI.
- Brown, Andrew R., Toby Gifford and Robert Davidson. 2015. "Techniques for Generative Melodies Inspired by Music Cognition." *Computer Music Journal* 39: 11-26.
- Burraston, Dave and Ernest Edmonds. 2005. "Cellular Automata in Generative Electronic Music and Sonic Art: A Historical and Technical Review." *Digital Creativity* 16: 165-185.
- Burraston, Dave. 2007. "Fundamental Insights on Complex Systems Arising from Generative Arts Practice." *Leonardo* 40: 372-373.
- Burgess, Jean and Joshua Green. 2009. *YouTube: Online Video and Participatory Culture*. Cambridge: Polity Press.
- Candy, Linda and Ernest A. Edmonds. 2002. *Explorations in Art and Technology*. Berlin: Springer.
- Carroll, Noël. 1996. *Theorizing the Moving Image*. Cambridge: Cambridge University Press.

- Cavell, Stanley. 1979. *The World Viewed: Reflections on the Ontology of Film*. Cambridge: Harvard University Press.
- Chaitin, Gregory. 2001. *Exploring Randomness*. London: Springer, 2001.
- Chambel, Teresa, Luís Correia, Jônatas Manzolli, Goncalo Dias Miguel, Nuno A.C. Henriques and Nuno Correia. 2007. "Creating Video Art with Evolutionary Algorithms." *Computer & Graphics* 31: 837-847.
- Chen, Fan, Damien Delannay and Christophe de Vleeschouwer. 2011. "An Autonomous Framework to Produce and Distribute Personalized Team-Sport Video Summaries: A Basket-Ball Case Study." *IEEE Transactions on Multimedia* 13: 1281-1394.
- Chion, Michel. 1994. *Audio-Vision: Sound on Screen*. New York: Columbia University Press.
- Chomsky, Noam. 1975. *Syntactic Structures*. Mouton: The Hague.
- Chorianopoulos, Konstantinos and George Lekakos. 2008. "Introduction to Social TV: Enhancing the Shared Experience with Interactive TV." *International Journal of Human-Computer Interaction* 24: 113-120.
- Clark, Andy. 2003. *Natural-Born Cyborgs: Minds, Technologies, and the Future of Human Intelligence*. Oxford: Oxford University Press.
- Combs, Allan and Mark Holland. 1990. *Synchronicity: Science, Myth, and the Trickster*. New York: Paragon House.
- Condry, John C. 1989. *The Psychology of Television*. Hillsdale: L. Erlbaum Assoc.
- Cook, David A. 2004. *A History of Narrative Film*. New York: Norton.
- Coveney, Peter and Roger Highfield. 1995. *Frontiers of Complexity: The Search for Order in a Chaotic World*. New York: Ballantine Books.
- Cowan, Michael. 2007. "The Heart Machine: 'Rhythm' and Body in Weimar Film and Fritz Lang's Metropolis." *Modernism/Modernity* 14: 225-248.
- Crimp, Douglas. 1981. "The End of Painting." *October* 16: 69-86.
- Cubitt, Sean. 2004. *The Cinema Effect*. Cambridge: MIT Press.
- Cutting, James E., Jordan E. DeLong and Christine E. Nothelfer. 2010. "Attention and the Evolution of Hollywood Film." *Psychological Science* 20: 1-8.
- Cutting, James E., Kaitlin L. Brunick, Jordan E. DeLong and Ayse Candan. 2011. "Quicker, Faster, Darker: Changes in Hollywood Film over 75 Years." *i-Perception* 2: 569-576.
- Dancyger, Ken. 2011. *The Technique of Film and Video Editing: History, Theory, and Practice*. Burlington: Elsevier/Focal Press.
- Danesi, Marcel. 2008. *Popular Culture: Introductory Perspectives*. Lanham: Rowman & Littlefield.
- Daval, Jean-Luc. 1982. *Photography: History of an Art*. Surrey: Macmillan.
- Davenport, Glorianna, Thomas Aguirre Smith and Natalio Pincever. 1991. "Cinematic Primitives for Multimedia." *IEEE Computer Graphics & Applications* 11: 67-74.
- Dawkins, Richard. 1976. *The Selfish Gene*. New York: Oxford University Press.
- Deery, June. 2003. "TV.com: Participatory Viewing on the Web." *Journal of Popular Culture* 37: 161-183.

- Dmytryk, Edward. 1984. *On Film Editing: An Introduction to the Art of Film Construction*. Boston: Focal Press.
- Dorin, Alan. 2013. "Chance and Complexity: Stochastic and Generative Processes in Art and Creativity." In *Proceedings of the Virtual Reality International Conference: Laval Virtual*, edited by Simon Richir, 1-8. New York: ACM.
- Edmond, Maura. 2012. "Here We Go Again: Music Videos after YouTube." *Television & New Media*: 1-16.
- Eigenfeldt, Arne, Miles Thorogood, Jim Bizzocchi, Philippe Pasquier and Tom Calvert. 2014. "Video, Music, and Sound Metacreation." In *xCoAx 2014: Proceedings of the Second Conference on Computation, Communication, Aesthetics and X*, edited by Miguel Carvalhais and Mario Verdicchio, 320-332. Porto: Universidade de Porto.
- Engelbart, Douglas. 2003. "Augmenting Human Intellect." In *The New Media Reader*, edited by Noah Wardrip-Fruin and Nick Montfort, 93-108. Cambridge: MIT Press.
- Eno, Brian. 2004. "The Studio as Compositional Tool." In *Audio Culture: Readings in Modern Music*, edited by Christoph Cox and Daniel Warner, 127-130. New York: Continuum.
- Fiske, John. 1987. *Television Culture*. London: Routledge.
- Flake, Gary William. 1998. *The Computational Beauty of Nature: Computer Explorations of Fractals, Chaos, Complex Systems, and Adaptation*. Cambridge: MIT Press.
- Flew, Terry. 2008. *New Media: An Introduction*. South Melbourne: Oxford University Press.
- Follett, Danielle. 2011. "Tout et N'importe Quoi: The Total Artwork and the Aesthetics of Chance." In *The Aesthetics of the Total Artwork – On Borders and Fragments*, edited by Anke Finger and Danielle Follett, 86-109. Baltimore: The Johns Hopkins University Press.
- Follett, Danielle and Anke Finger. 2011. "Dynamiting the Gesamtkunstwerk – An Introduction to the Aesthetics of the Total Artwork", in *The Aesthetics of the Total Artwork – On Borders and Fragments*, edited by Anke Finger and Danielle Follett, 1-25. Baltimore: The Johns Hopkins University Press.
- Foote, Jonathan, Matthew Cooper and Andreas Grgensohn. 2002 "Creating Music Videos Using Automatic Media Analysis." In *Proceedings of the 10th ACM International Conference on Multimedia*, 553-560.
- Frey, Carl Benedikt and Michael A. Osborne. 2013. "The Future of Employment: How Susceptible Are Jobs to Computerisation?", *unpublished manuscript*.
- Frith, Simon, Andrew Goodwin and Lawrence Grossberg. 1993. *Sound & Vision: The Music Video Reader*. New York: Routledge.
- Gabrielli, Giulia. 2010. "An Analysis of the Relation Between Music and Image. The Contribution of Michel Gondry." In *Rewind Play Fast Forward. The Past, Present and Future of the Music Video*, edited by Henry Keazor and Thorsten Wübbena, 89-109. Bielefeld: transcript.
- Galanter, Philip. 2003. "What is Generative Art? Complexity Theory as a Context for Art Theory." In *International Conference on Generative Art*. Milan.

- . 2008. "Complexism and the Role of Evolutionary Art." In *The Art of Artificial Evolution*, edited by Juan Romero and Penousal Machado, 311-332. Berlin: Springer.
- Girgensohn, Andreas, John Boreczky, Patrick Chiu, John Doherty, Jonathan Foote, Gene Golovchinsky, Shingo Uchihashi and Lynn Wilcox. 2000. "A Semi-Automatic Approach to Home Video Editing." In *Proceedings of UIST*, 81-89.
- Goldreich, Oded. 2008. *Computational Complexity: A Conceptual Perspective*. New York: Cambridge University Press.
- Goodwin, Andrew. 1992. *Dancing in the Distraction Factory: Music Television and Popular Culture*. Minneapolis: University of Minnesota Press.
- Gow, Joe. 1992. "Music Video as Communication: Popular Formulas and Emerging Genres." *Journal of Popular Culture* 26: 41-70.
- Griffiths, Paul. 1981. *Cage*. London: Oxford University Press.
- Grinstead, Charles M. 1997. *Introduction to Probability*. Providence: American Mathematical Society.
- Gunning, Tom. 1994. DW Griffith and the Origins of American Narrative Film: The Early Years at Biograph. Illinois: University of Illinois Press.
- Halpern, Megan and Lee Humphreys. 2014. "Iphoneography as an Emergent Art World." *New Media & Society*: 1-20.
- Hammoud, Riad I. 2006. *Interactive Video: Algorithms and Technologies*. Berlin: Springer.
- Heidegger, Martin. 1977. *The Question Concerning Technology and Other Essays*. New York: Harper & Row.
- Helvey T.C. 1971. *The Age of Information: an Interdisciplinary Survey of Cybernetics*. Englewood Cliffs: Educational Technology Publications.
- Herzog, Amy. 2007. "Illustrating Music: The Impossible Embodiments of the Jukebox Film." In *Medium Cool: Music Videos from Soundies to Cellphones*, edited by Roger Beebe and Jason Middleton, 30-58. Durham: Duke University Press.
- Hickman, Larry A. 1990. *John Dewey's Pragmatic Technology*. Bloomington: Indiana University Press.
- Holdstein, Deborah H. 1984. "Music Video: Messages and Structures." *Jump Cut* 29: 1-13.
- Hu, Paul Jen-Hwa, Theodore H.K. Clark and Will W. Ma. 2003. "Examining Technology Acceptance by School Teachers: A Longitudinal Study." *Information & Management* 41: 227-241.
- Hua, Xian-Sheng, Lie Lu and Hong-Jiang Zhan. 2004. "Optimization-Based Automated Home Video Editing System." *IEEE Transactions on Circuits and Systems for Video Technology* 14: 572-583.
- Irwin, Kathleen. 2011. "Staging the Internet: Representation (Bodies, Memories) and Digital Audiences." *Canadian Theatre Review* 148: 54-60.
- Jenkins, Henry. 2006. *Convergence Culture: Where Old and New Media Collide*. New York: New York University Press.
- Jenkins, Henry, Sam Ford and Joshua Green. 2013. *Spreadable Media: Creating Value and Meaning in a Networked Culture*. New York: New York University Press.
- Jennings, Nicholas R., Katia Sycara and Michael Wooldridge. 1998. "A Roadmap of Agent Research and Development." *Autonomous Agents and Multi-Agent Systems* 1: 7-38.

- Jhally, Sut. 1995. "Image-Based Culture: Advertising and Popular Culture." In *Gender, Race, and Class in Media*, edited by Gail Dines and Jean M. Humez, 77-87. Thousand Oaks: Sage Publications.
- John, Nicholas A. 2012. "Sharing and Web 2.0: The Emergence of a Keyword." *New Media Society* 15: 167-182.
- Jones, Steve. 1988. "Cohesive but not Coherent: Music Videos, Narrative and Culture." *Popular Music and Society* 12: 15-29.
- Jones, Steve. 2005. "MTV: The Medium Was the Message." *Critical Studies in Media Communication* 22: 83-88.
- Jung, Carl Gustav. 1967. *Memories, Dreams, Reflections*. London: Collins.
- Kaplan, E. Ann. 1988. *Rocking Around the Clock*. London: Routledge.
- Keazor, Henry and Thorsten Wübbena. 2007. *Video Thrills the Radio Star: Musikvideos: Geschichte, Themen, Analysen*. Bielefeld: transcript.
- . 2010. "Introduction." In *Rewind, Play, Fast Forward. The Past, Present and Future of the Music Video*, edited by Henry Keazor and Thorsten Wübbena, 7-31. Bielefeld: transcript.
- Keen, Andrew. 2007. *The Cult of the Amateur: How Today's Internet is Killing Our Culture*. New York: Doubleday/Currency.
- Keith, Susan. 2014. "Back to the 1990s? Comparing the Discourses of 20th- and 21st-Century Digital Image Ethics Debates." *Visual Communication Quarterly* 21: 61-71.
- Kern, Angela. 2011. "Wir Wirkt ein Kleines Bild – Betrachtung aus der Sicht der Medienproduktion", in *Zur Ästhetischen Umsetzung von Musikvideo im Kontext von Handhelds*, edited by Henry Keazor, Hans. W. Giessen and Thorsten Wübbena, 121-137. Heidelberg: Ruprechts-Karls-Universität.
- Kim, Pyungho. 2001. "New Media, Old Ideas: The Organizing Ideology of Interactive TV." *Journal of Communication Inquiry* 25: 72-88.
- Knuth, Donald Ervin. 1969. *The Art of Computer Programming*. Reading: Addison-Wesley Pub. Co.
- Kosmopoulos, D. I., A. Doulamis, A. Makris, N. Doulamis, S. Chatzis and E. Middleton. 2009. "Vision-Based Production of Personalized Video." *Signal Processing: Image Communication* 24: 158-176.
- Kubicek, Radek, Pavel Zak, Pavel Zemcik and Adam Herout. 2008. "Automatic Video Editing for Multimodal Meetings", in *ICCVG'08*, 1-12.
- Kuleshov, Lev Vladimirovich. "The Principles of Montage." In *Kuleshov on Film. Writings of Lev Kuleshov*, 183-195. Berkeley: University of California Press.
- Lashkari, Yezdi, Max Metral and Pattie Maes. 1998. "Collaborative Interface Agents." In *Readings in Agents*, edited by Michael N. Huhns and Munindar Paul Singh, 111-116. San Francisco: Morgan Kaufmann.
- Lavin, Sylvia. 2011. "Scenes from the Suburbs: The Wilderness Downtown." *Journal of the Society of Architectural Historians* 70: 398-400.
- Le Grice, Malcolm. 1977. *Abstract Film and Beyond*. London: Studio Vista.
- Lee, Barbara and Robert S. Lee. 1995. "How and Why People Watch TV: Implications for the Future of Interactive Television." *Journal of Advertising Research* 35: 9-18.

- Leibniz, Gottfried Wilhelm. 1994. *Writings on China*. Chicago: Open Court.
- Leggett, Mike and Zafa Bilda. 2008. "Exploring Design Options for Interactive Video with the Mnemovie Hypervideo System." *Design Studies* 29: 587-602.
- LeWitt, Sol. 1967. "Paragraphs on Conceptual Art." *Artform* 5: 79-83.
- Licklider, Joseph Carl Robnett. 2003. "Man-Computer Symbiosis." In *The New Media Reader*, edited by Noah Wardrip-Fruin and Nick Montfort, 73-82. Cambridge: MIT Press.
- Lima, Edirlei Soares, Bruno Feijó, Antonio L Furtado and Cesar Pozzer. 2012. "Automatic Video Editing for Video-Based Interactive Storytelling." In *Proceedings of the International Conference on Multimedia & Expo*, 806-811.
- Lister, Martin, Jon Dovey, Seth Giddings, Grant Iain, Kieran Kelly. 2009. *New Media: A Critical Introduction*. New York: Routledge.
- Lotz, Amanda. 2007. *Television Will Be Revolutionized*. New York: New York University Press.
- Lu, Peter J. and Paul J. Steinhardt. 2007. "Decagonal and Quasi-Crystalline Tilings in Medieval Islamic Architecture." *Science* 315: 1106-1110.
- Lunenfeld, Peter. 2004. "The Myths of Interactive Cinema." In *Narrative Across Media: The Language of Storytelling*, edited by Marie-Laure Ryan, 377-390. Lincoln: University of Nebraska Press.
- Lynch, Joan D. 1984. "Muisic Videos: From Performance to Dada-Surrealism." *Journal of Popular Culture* 18: 53-57.
- Mackay, Wendy E. and Glorianna Davenport. (1989) "Virtual Video Editing in Interactive Multimedia Applications." *Communications of the ACM* 32: 802-810.
- Manovich, Lev. 2001. *The Language of New Media*. Cambridge: MIT Press.
- . 2003. "New Media From Borges to HTML." In *The New Media Reader*, edited by Noah Wardrip-Fruin and Nick Montfort, 13-25. Cambridge: MIT Press.
- . 2013. *Software Takes Command*. New York: Bloomsbury Academic.
- Manovich, Lev and Andreas Kratky. 2005. *Soft Cinema: Navigating the Database*. Cambridge: MIT Press.
- McAlpine, Kenneth, Eduardo Miranda and Stuart Hoggar. 1999. "Making Music with Algorithms: A Case-Study System." *Computer Music Journal* 23: 19-30.
- McCorduck, Pamela. 1991. *Aaron's Code: Meta-Art, Artificial Intelligence, and the Work of Harold Cohen*. New York: W.H. Freeman & Co.
- McCormack, Jon and Alan Dorin. 2001. "Art, Emergence and the Computational Sublime." In *Second Iteration: A Conference on Generative Systems in the Electronic Arts*, edited by Alan Dorin, 67-81. Melbourne: CEMA.
- McCormack, Jon and Mark d'Inverno. 2012. *Computers and Creativity*. Berlin: Springer.
- McCormack, Jon, Oliver Bown, Alan Dorin, Jonathan McCabe, Gordon Monro and Mitchell Whitelaw. 2014. "Ten Questions Concerning Generative Computer Art." *Leonardo* 47: 135-141.
- McLuhan, Marshall. 2001. *Understanding Media: The Extensions of Man*. London: Routledge.
- . 2004. "Visual and Acoustic Space." In *Audio Culture: Readings in Modern Music*, edited by Christoph Cox and Daniel Warner, 67-72. New York: Continuum.

- Meixner, Britta, Katarzyna Matusik, Christoph Grill and Harald Kosch. 2014. "Towards an Easy to Use Authoring Tool for Interactive Non-Linear Video." *Multimedia Tools and Applications* 70: 1251-1276.
- Metz, Christian. 1974. *Film Language: A Semiotics of the Cinema*. New York: Oxford University Press.
- Mjos, Ole J. 2012. *Music, Social Media, and Global Mobility*. New York: Routledge.
- Mlodinow, Leonard. 2008. *The Drunkard's Walk: How Randomness Rules Our Lives*. New York: Pantheon Books.
- Monaco, James. 1981. *How to Read a Film: Movies, Media, Multimedia*. New York: Oxford University Press.
- Moore, Omar Khayyam. 1957. "Divination – A New Perspective." *American Anthropologist* 59: 69-74.
- Morse, Margaret. 1986. "Postsynchronizing Rock Music and Television." *Journal of Communication Inquiry* 10: 15-28.
- Murch, Walter. 2001. *In the Blink of an Eye: A Perspective on Film Editing*. Los Angeles: Silman-James Press.
- Nack, Frank Michael. 1996. "AUTEUR: The Application of Video Semantics and Theme Representation for Automated Film Editing." PhD diss., Lancaster University.
- Nake, Frieder. 2010. "Paragraphs on Computer Art, Past and Present." In *CAT 2010: Ideas Before Their Time*, edited by Nick Lambert, Jeremy Gardiner and Francesca Franco. 55-63. Swinton: British Computer Society.
- . 2012. "Construction and Intuition: Creativity in Early Computer Art." In *Computers and Creativity*, edited Jon McCormack and Mark d'Inverno, 61-94. Berlin: Springer.
- Napoli, Philip M. 2014. "Automated Media: An Institutional Theory Perspective on Algorithmic Media Production and Consumption." *Communication Theory* 24: 340-360.
- Negroponte, Nicholas. 1995. *Being Digital*. New York: Knopf.
- Nelson, Theodor H. 1987. "Dream Machines." In *Computer Lib/Dream Machines*, edited by Theodor H. Nelson, nP. Redmon: Tempus Books of Microsoft Press.
- Nelson, Robin. 1997. *TV Drama in Transition: Forms, Values, and Cultural Change*. New York: St. Martin's Press.
- Nichols, Bill. 1975. "Style, Grammar, and the Movies." *Film Quarterly* 28: 33-49.
- Nierhaus, Gerhard. 2009. *Algorithmic Composition: Paradigms of Automated Music Generation*. Vienna: Springer.
- Nithya, C. and M. Privadharsini. 2013. "A Framework for Automatic Script-to-Movie Production." *International Journal of Advanced Research in Computer and Communication Engineering* 2: 4523-4526.
- Noble, David F. 1998. "Digital Diploma Mills: The Automation of Higher Education." *Science as Culture* 7: 355-368.
- Nye, David E. 2006. *Technology Matters: Questions to Live With*. Cambridge: MIT Press.
- O'Hear, Anthony. 1995. "Art and Technology: An Old Tension." *Royal Institute of Philosophy Supplement* 38: 143-158.

- Oldham, Gabriella. 1992. *First Cut: Conversations with Film Editors*. Berkeley: University of California Press.
- . 2012. *First Cut 2: More Conversations with Film Editors*. Berkeley: University of California Press.
- Ondaatje, Michael. 2002. *The Conversations: Walter Murch and the Art of Editing Film*. London: Bloomsbury.
- Paik, Nam June. 1993. *Eine DATA base*. Stuttgart: Edition Cantz.
- Peverini, Paolo. 2010. "The Aesthetics of Music Videos: An Open Debate." In *Rewind, Play, Fast Forward. The Past, Present and Future of the Music Video*, edited by Henry Keazor and Thorsten Wübbena, 135-153. Bielefeld: transcript.
- Piacenza, Alberto, Fabrizio Guerrini, Nicola Adami, Riccardo Leonardi, Jonathan Teutenberg, Julie Porteous and Marc Cavazza. 2011. "Generating Story Variants with Constrained Video Recombination." In *MM '11 Proceedings of the 19th ACM International Conference on Multimedia*, 223-232. New York: ACM.
- Pool, Ithiel de Sola. 1983. *Technologies of Freedom*. Cambridge: Belknap Press.
- Prince, Gerald. 2003. *A Dictionary of Narratology*. Nebraska: University of Nebraska Press.
- Rabiger, Michael. 1989. *Directing – Film Techniques and Aesthetics*. Boston: Focal Press.
- Railton, Diane and Paul Watson. 2011. *Music Video and the Politics of Representation*. Edinburgh: Edinburgh University Press.
- Reisz, Karel. 2010. *The Technique of Film Editing*. Amsterdam: Elsevier.
- Rieser, Martin and Andrea Zapp. 2002. "An Age of Narrative Chaos?" In *New Screen Media: Cinema/Art/Narrative*, edited by Martin Rieser and Andrea Zapp, xxv-xxviii. London: British Film Institute.
- Rimmon-Kenan, Shlomith. 1989. *Narrative Fiction: Contemporary Poetics*. London: Routledge.
- Robbe-Grillet, Alain and Bruce Morrisette. 1977. "Order and Disorder in Film and Fiction." *Critical Inquiry* 4: 1-20.
- Rodowick, David Norman. 2007. *The Virtual Life of Film*. Cambridge: Harvard University Press.
- Romero, Juan and Penousal Machado. 2008. *The Art of Artificial Evolution*. Berlin: Springer.
- Schmitt, Thomas. 2010. "The Genealogy of the Clip Culture." In *Rewind, Play, Fast Forward. The Past, Present and Future of the Music Video*, edited by Henry Keazor and Thorsten Wübbena, 41-57. Bielefeld: transcript.
- Shanken, Edward A. 2002. "Art in the Information Age: Technology and Conceptual Art." *Leonardo* 35: 433-438.
- Shifman, Limor. 2014. *Memes in Digital Culture*. Cambridge: MIT Press.
- Shim, Doobo. 2006. "Hybridity and the Rise of Korean Popular Culture in Asia." *Media Culture Society* 28: 25-44.
- Shneiderman, Ben. 2000. "Creating Creativity: User Interfaces for Supporting Innovation." *ACM Transactions on Computer-Human Interaction* 7: 114.138.

- Shneiderman, Ben, Gerhard Fischer, Mary Czerwinsky, Mitch Resnick, Brad Myers, Linda Candy, Ernest Edmonds, Mike Eisenberg, Elisa Giaccardi, Tom Hewett, Pamela Jennings, Bill Kules, Kumiyo Nakakoji, Jay Nunamaker, Randy Pausch, Ted Selker, Elisabeth Sylvan and Michael Terry. 2006. "Creativity Support Tools: Report from a U.S. National Science Foundation Sponsored Workshop." *International Journal of Human-Computer Interaction* 20: 61-77.
- Sibilla, Gianni. 2010. "It's the End of Music Videos as We Know Them (But We Feel Fine)." In *Rewind, Play, Fast Forward. The Past, Present and Future of the Music Video*, edited by Henry Keazor and Thorsten Wübbena, 225-232. Bielefeld: transcript.
- Smith, David E., Jeremy Frank, and Ari K. Jónsson. 2000. "Bridging the Gap Between Planning and Scheduling." *The Knowledge Engineering Review* 15: 47-83.
- Snickars, Pelle and Patrick Vonderau. 2009. *The YouTube Reader*. Stockholm: National Library of Sweden.
- Snoek, Cees G. M. and Marcel Worring. 2005. "Multimodal Video Indexing: A Review of the State-of-the-Art." *Multimedia Tools and Applications* 25: 5-35.
- Sorensen, Andrew C. and Andrew R. Brown. 2009. "Interacting with Generative Music through Live Coding." *Contemporary Music Review* 28: 17-29.
- Stam, Robert, Robert Burgoyne and Sandy Flitterman-Lewis. 1992. *New Vocabularies in Film Semiotics: Structuralism, Post-Structuralism, and Beyond*. London: Routledge.
- Strachan, Rob. 2006. "Music Video and Genre: Structure, Context, and Commerce." In *Music and Manipulation: On the Social Uses and Social Control of Music*, edited by Steven Brown and Ulrik Volgsten, 187-206. New York: Berghahn Books.
- Straw, Will. 1993. "Popular Music and Postmodernism in the 1980s." In *Sound & Vision: The Music Video Reader*, edited by Simon Frith, Andrew Goodwin and Lawrence Grossberg, 3-21. New York: Routledge.
- Szilas, Nicolas. 1999. "Interactive Drama on Computer: Beyond Linear Narrative", in *Proceedings of the AAAI Fall Symposium: Narrative Intelligence Technical Report*, 150-156. Menlo Park: AAAI Press.
- Terpstra, P. 1961. *The Graphic Work of M.C. Escher*. London: Oldbourne.
- Todorov, Tzvetan. 1971. "The 2 Principles of Narrative." *Diacritics* 1: 37-44.
- Uricchio, William. 2009. "The Future of a Medium Once Known as Television." In *The YouTube Reader*, edited by Pelle Snickars and Patrick Vonderau, 24-39. Stockholm: National Library of Sweden.
- Vernallis, Carol. 2001. "The Kindest Cut: Functions and Meanings of Music Video Editing." *Screen* 42: 21-48.
- . 2004. *Experiencing Music Video: Aesthetics and Cultural Context*. New York: Columbia University Press.
- . 2007. "Strange People, Weird Objects: The Nature of Narrativity, Character, and Editing in Music Videos", in *Medium Cool: Music Videos from Soundies to Cellphones*, edited by Roger Beebe and Jason Middleton , 111-151. Durham: Duke University Press.

- . 2010. "Music Video and YouTube: New Aesthetics and Generic Transformations", in *Rewind, Play, Fast Forward. The Past, Present and Future of the Music Video*, edited by Henry Keazor and Thorsten Wübbena, 233-261. Bielefeld: transcript.
- . 2013. *Unruly Media: YouTube, Music Video, and the New Digital Cinema*. New York: Oxford University Press.
- Viola, Bill. 2003. "Will There Be Condominiums in Data Space?" In *The New Media Reader*, edited by Noah Wardrip-Fruin and Nick Montfort, 463-470. Cambridge: MIT Press.
- Vorderer, Peter. 2000. "Entertainment, Suspense, and Interactivity" In *Moving Images, Culture and the Mind*, edited by Ib Bjondeberg, 65-83. Luton: University of Luton Press.
- Wang, Jinjun, Changsheng Xu, Engsiong Chng, Lingyu Duan, Kongwah Wan and Qi Tian. 2005. "Automatic Generation of Personalized Music Sports Video." In *Proceedings of ACM MultiMedia '05*: 7735-744.
- Wang, Jinjun, Engsiong Chng, Chengshen Xu, Hanqiq Lu and Qi Tian. 2007. "Generation of Personalized Music Sports Video Using Multimodal Cues." *IEEE Transactions Multimedia* 9: 576-588.
- Wardrip-Fruin, Noah and Nick Montfort. 2003. *The New Media Reader*. Cambridge: MIT Press.
- Weizenbaum, Joseph. 1984. *Computer Power and Human Reason: From Judgment to Calculation*. Harmondsworth: Penguin.
- Welsby, Chris. 2006. "Films and Installations – A Systems View of Nature." In *Experimental Film and Video: An Anthology*, edited by Jackie Hatfield, 26-35. Eastleigh: John Libbey.
- Whitelaw, Mitchell. 2004. *Metacreation: Art and Artificial Life*. Cambridge: MIT Press.
- Wiggins, Bradley E. and Bret G. Bowers. 2014. "Memes as Genre: A Structurational Analysis of the Memescape." *New Media & Society* 1: 1-21.
- Willemen, Paul. 2002. "Reflections on 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.
- Williams, Kevin. 2003. *Why I [Still] Want My MTV*. Cresskill: Hampton Press.
- Wilscher, Gundula. 2011. "Mobile Music: App(2)Date. Vier Jahre Austrian Mobile Music Day", in *Zur Ästhetischen Umsetzung von Musikvideo im Kontext von Handhelds*, edited by Henry Keazor, Hans. W. Giessen and Thorsten Wübbena, 121-137. Heidelberg: Ruprechts-Karls-Universität.
- Wolfe, Arnold S. 1983. "Rock on Cable: On MTV: Music Television, the First Video Music Channel." *Popular Music and Society* 9: 41-50.
- Wooldridge, Michael and Nicholas R. Jennings. 1995. "Intelligent Agents: Theory and Practice." *The Knowledge Engineering Review* 10: 115-152.
- Zettl, Herbert. 2014. *Sight Sound Motion: Applied Media Aesthetics*. Boston: Wadsworth.

Appendix A

1	Artist
2	Title
3	Genre 1
4	Genre 2
5	Genre 3
6	Linearity
7	Follows song structure
8	Rhythmic editing
9	Narrative level
10	Narrative (content type)
11	Dance (content type)
12	Performance (content type)
13	Abstract/concept (content type)
14	Segments
15	Lip-synch
16	Lyric references
17	Beats per minute
18	Shots per minute

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
Angus & Julia Stone	Big Jet Plane	Folk	Acoustic	Indie	semi-linear	weak	weak	2	y	n	n	3	y	weak	110	9.83	
Little Red	Rock It	Indie Pop	Rock	semi-linear	medium	medium	3	n	y	y	n	4	y	medium	107	25.67	
Ouest le swimming pool	Dance the way I feel	Electronic	Indie	Synthpop	non-linear	strong	6	n	y	y	n	6	y	weak	118	effects	
Birds of Tokyo	Plans	Alternative Rock	Rock	Alternative	non-linear	weak	weak	6	n	n	n	y	0	y	weak	103	23.67
Adrian Lux	Teenage Crime	House	Electronic	linear	medium	weak	1	y	n	n	3	n	weak	128	44.07		
Cee-Lo Green	Fuck you	Soul	Funk	RnB	linear	medium	strong	1	y	n	n	5	y	strong	127	24.72	
The Wombats	Tokyo	Indie	Indie Rock	Britpop	semi-linear	medium	strong	1	y	y	n	0	y	medium	145	38.52	
Art vs Science	Magic Fountain	Electronic	Indie	New Rave	linear	medium	strong	1	y	y	n	4	n	weak	127	20.09	
Mark Ronson & the Business Int'l	Somebody to love me	Electronic	Indie	Hip Hop	semi-linear	strong	weak	2	y	n	n	8	y	weak	118	9.01	
Draft	Rapunzel	Hip Hop	Rap	linear	weak	weak	1	y	n	n	0	n	strong	120	23.17		
Sia	Clap your hands	Indie	Downtempo	Trip-Hop	non-linear	weak	weak	6	n	n	n	y	0	medium	120	18.38	
Kanye West	Runaway	Hip Hop	Rap	RnB	linear	weak	medium	6	n	y	y	n	3	y	medium	90	7.32
Duck Sauce	Barbra Streisand	Dance	Disco	Funk	non-linear	weak	weak	4	n	y	y	n	0	y	weak	128	49.29
Mark Ronson & the Business Int'l	Bang Bang	Electronic	Indie	Hip Hop	non-linear	strong	strong	6	n	y	y	n	6	y	strong	93	44.76
Cloud Control	There's nothing in the water we can't fight	Folk rock	Indie folk		non-linear	strong	medium	6	n	n	y	y	7	y	weak	100	15
Flight Facilities	Crave you	Pop	Downtempo	Nu Disco	semi-linear	medium	weak	2	y	n	n	y	3	y	weak	120	28.31
Washington	Sunday Best	Indie Pop	Rock	semi-linear	weak	medium	6	n	y	y	n	3	y	weak	165	32.8	
Two Door Cinema Club	Undercover	Indie Rock	Indie Pop	Rock	non-linear	weak	strong	6	n	n	y	y	4	y	medium	160	42.86
Children	Jelly Legs	Indie Rock	Indie	Rock	non-linear	weak	medium	6	n	n	y	0	y	weak	157	40.55	

Collide											
Bliss n Eso	Addicted	Hip Hop	Rap	linear	weak	strong	7	n	n	y	0
Sparkadia	Talking like I'm falling down stairs	Indie	Indie Pop	Pop	non-linear	medium	strong	6	n	n	y
Goye	Eyes wide open	Electronic	Indie	Alternative	semi-linear	weak	medium	3	y	n	0
Florence & The Machine feat Dizzee Rascal	You got the dirtree love (live)	Indie	Indie Pop	Hip Hop	linear	weak	weak	5	n	n	0
Radar Detector	Darwin Deez	Indie	Indie Pop	Electronic	non-linear	weak	strong	6	n	n	y
It can wait	Ily	Hip Hop	Rap	non-linear	medium	medium	6	n	n	y	0
Yeasayer	O.N.E.	Indie	Indie Rock	Folk	semi-linear	medium	2	y	y	n	0
The National	Bloodbuzz Ohio	Indie	Indie Rock	Alternative	non-linear	weak	weak	6	n	n	y
Foster the People	Pumped up kicks	Indie	Indie Pop	Alternative	non-linear	medium	weak	4	n	n	y
Tame Impala	Solitude is bliss	Psychedelic Rock	Indie Rock	Rock	semi-linear	strong	medium	2	y	n	4
The Naked and Famous	Punching in a dream	Indie	Electronic	Electro Pop	non-linear	strong	medium	3	n	n	y
Mark Ronson & the Business Int'l	The Bike Song	Electronic	Indie	Hip Hop	linear	strong	strong	2	y	n	n
Chiddy Bang	Opposite of adults	Hip Hop	Rap	Electronic	non-linear	weak	weak	6	n	n	y
Gorillaz feat Daley	Doncamatic	Alternative	Electronic	Hip Hop	semi-linear	weak	medium	1	y	n	0
The Naked and Famous	Young Blood	Indie	Electronic	Electro Pop	non-linear	medium	4	n	n	y	5
The John Butler Trio	Revolution	Acoustic	Rock	Alternative	non-linear	weak	medium	6	n	n	y
Gyroscope	Baby I'm getting better	Rock	Alternative	Rock	semi-linear	weak	weak	3	n	n	0
Bliss n Eso	Down by the river	Hip Hop	Rap	non-linear	weak	medium	4	n	n	y	0
Gorillaz	On Melancholy Hill	Alternative	Electronic	Hip Hop	semi-linear	medium	weak	2	y	n	3
Yolande Be Cool	We no speak americano	Electronic	Dance	House	linear	weak	weak	2	y	n	0

Crystal Castles	Baptism	Electronic	New Rave	non-linear	strong	strong	6	n	y	y	6	y	weak	120	12.07		
Boy & Bear	Rabbit Song	Folk	Indie	Indie Folk	semi-linear	weak	weak	1	y	n	n	0	n	weak	104	28.39	
Bag Raiders	Way back home	Electronic	Dance	Funk	semi-linear	medium	strong	7	n	n	y	0	n	medium	126	1.11	
Birds of Tokyo	Wild at heart	Alternative Rock	Rock	Alternative	non-linear	weak	medium	3	n	n	y	0	y	weak	100	55.23	
Pendulum	Witchcraft	drum and electronic bass	electronic	semi-linear	medium	strong	3	n	n	y	y	3	y	medium	174	51.83	
The Jezebels	Easy to love	Indie	Rock	semi-linear	weak	weak	5	n	n	y	n	0	y	weak	157 (?)	27.91	
Hot Chip	One life stand	Electronic	Indie	Electropop	non-linear	weak	weak	6	n	n	y	0	y	weak	120	14.85	
Yeasayer	Ambling Alp	Indie	Indie Rock	Folk	non-linear	medium	strong	3	n	n	y	0	y	weak	144	48	
The John Steel Singers	Overpass	Indie Pop	Trout Rock	Pop	linear	weak	medium	2	y	n	n	0	y	medium	131	33.96	
Miami Horror	Holidays	Electronic	Synthpop	Indie	semi-linear	medium	strong	2	y	n	n	y	3	n	weak	116	39.38
Vampire Weekend	Giving up the gun	Indie	Indie Rock	Indie Pop	semi-linear	weak	medium	2	y	n	y	y	5	y	weak	134	26.73
Example	Kickstarts	Hip Hop	Electronic	Rap	non-linear	medium	strong	3	n	y	n	y	8	y	strong	126	44.43
Arcade Fire	The Suburbs	Indie	Indie Rock	Alternative	linear	weak	weak	2	y	n	n	0	n	medium	117	18.67	
Washington	Rich Kids	Indie Pop	Rock	semi-linear	medium	medium	3	n	y	n	0	y	medium	182	38.7		
Children Collide	My Eagle	Indie Rock	Indie	Rock	linear	weak	medium	5	n	n	y	n	0	y	weak	140	27.72
Ou est le swimming pool	Jackson's Last Stand	Electronic	Indie	Synthpop	semi-linear	weak	weak	3	n	n	y	y	0	y	weak	120	44.52
Angus & Julia Stone	Hold On	Folk	Acoustic	Indie	semi-linear	weak	weak	3	n	n	y	0	n	weak	123	12.9	
Arcade Fire	Ready to Start	Indie	Indie Rock	Alternative	semi-linear	weak	weak	5	n	n	y	n	0	y	weak	99	48.28
Gypsy & The Cat	Jona Vark	Electronic	Indie	Electro Pop	semi-linear	medium	weak	6	n	n	y	y	2	y	weak	125	43.18
Dead Letter Circus	One Step Rock	Alternative Rock	Progressive Rock	Progressive	semi-linear	weak	weak	3	n	n	y	y	0	y	weak	105	28.1
Cold War Kids	Audience	Indie	Indie Rock	Alternative	semi-linear	medium	medium	2	y	n	y	y	4	y	weak	174	46.48
Vampire Weekend	Holiday	Indie	Indie Rock	Indie Pop	non-linear	weak	medium	3	n	n	y	0	y	weak	155	54.35	
Pendulum	Watercolour	drum and bass	electronic	semi-linear	medium	medium	3	y	y	3	y	y	3	y	weak	174	40.75

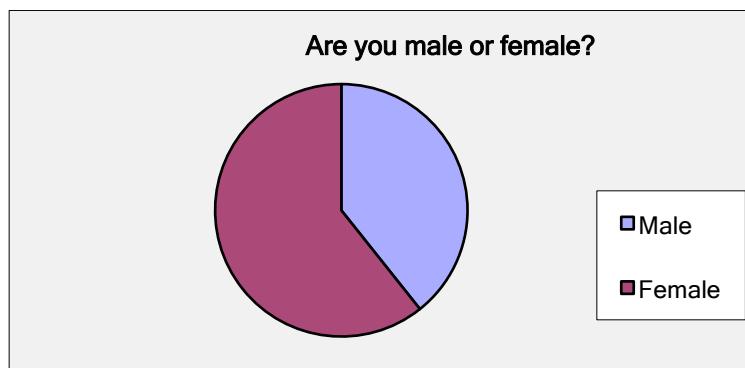
Groove Armada	Paper Romance	Electronic	Chillout	House	semi-linear	medium	2	y	n	y	n	0	y	weak	123	28.78	
Gypsy & The Cat	The Piper's Song	Electronic	Indie	Electro Pop	linear	weak	weak	5	n	n	y	n	0	y	weak	100	12.05
Two Door Cinema Club	I can talk	Indie Rock	Indie Pop	Rock	non-linear	strong	medium	6	n	n	y	y	6	y	weak	163	24.71
Tame Impala	Lucidity	Psychedelic Rock	Indie Rock	Rock	linear	weak	strong	7	n	n	n	y	0	n	weak	147	0
Hungry Kids of Hungary	Coming around	Indie	Indie Pop	Indie Rock	semi-linear	weak	weak	2	y	n	y	n	0	y	weak	179	23.72
Kings of Leon	Radioactive	Rock	Indie	Indie Rock	semi-linear	medium	medium	3	n	n	y	y	3	y	weak	108	40.17
Big Boi	Shutterbug	Hip Hop	Rap		non-linear	medium	medium	6	n	n	y	y	0	y	medium	98	54.94
Gorillaz	Stylo	Alternative	Electronic	Hip Hop	linear	weak	weak	1	y	n	n	n	0	y	weak	100	23.92
Little Red	Slow Motion	Indie Pop	Rock		semi-linear	weak	weak	6	n	n	y	y	0	y	weak	119	19.71
The Black Keys	Howlin for you	Blues Rock	Blues	Rock	linear	weak	weak	1	y	n	n	y	0	y	weak	132	37.5
Claxons	Echoes	New Rave	Indie	Electronic	semi-linear	strong	medium	6	n	n	y	y	5	y	weak	125	15.12
The Black Keys	Tighten Up	Blues Rock	Blues	Rock	linear	weak	weak	1	y	n	n	n	0	y	weak	109	35.9
Washington	The hardest part	Indie Pop	Rock		semi-linear	weak	medium	3	n	n	y	y	0	y	medium	130	31.33
Hot Chip	I feel better	Electronic	Indie	Electropop	linear	weak	weak	1	y	n	y	n	0	y	weak	126	34.57
Evil Eddie	Queensland	Hip Hop			semi-linear	medium	weak	2	y	n	n	y	0	y	medium	92	27.01
Birds of Tokyo	The saddest thing I know	Alternative Rock	Alternative	Post-Punk	non-linear	medium	weak	3	n	n	y	y	0	y	weak	100	20.1
Interpol	Barriade	Indie Rock			non-linear	strong	weak	6	n	n	y	y	0	y	weak	127	19.9
Art vs Science	Finally see our way	Electronic	Indie	New Rave	linear	weak	medium	1	y	n	n	n	0	n	weak	117	27.36
The Bedroom Philosopher	Northcote	Acoustic	Folk	Comedy	semi-linear	strong	weak	2	y	n	n	y	8	y	strong	141	37.18
The National	Anyone's Ghost	Indie	Indie Rock	Alternative	semi-linear	weak	medium	3	n	n	y	n	0	y	weak	122	25.58
Xavier Rudd	Time to smile	Acoustic	Folk	Singer-Songwriter	non-linear	weak	weak	4	n	n	y	y	0	y	weak	131	33.93
Broken Bells	The High Road	Indie	Experimental	Indie Pop	semi-linear	weak	weak	2	y	n	n	y	7	n	weak	160	27.79
Jonsi	Go Do	Ambient	Experimental	Indie	non-linear	weak	strong	6	n	n	n	y	3	y	weak	140	70.58
Parkway Drive	Sleepwalker	Hardcore	Metalcore	Melodic Metalcore	semi-linear	weak	weak	2	y	n	n	y	0	n	weak	167	28.19

Foals	Spanish Sahara	Indie	Math Rock	Indie Rock	semi-linear	medium	weak	3	y	n	y	3	y	weak	131	11.32	
Dead Letter Circus	Big Rock	Alternative Rock	Progressive Rock	Progressive Rock	linear	weak	weak	3	y	n	y	6	y	weak	123	20.63348416	
Muse	Neutron Star Collision	Alternative Rock	Rock	Alternative	semi-linear	weak	weak	2	y	n	y	n	0	y	weak	127	40.56680162

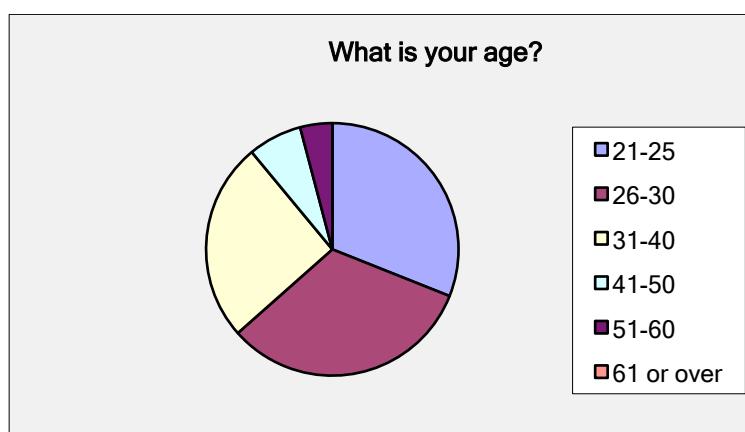
APPENDIX B

Ethics GU Ref No: QCA/33/12/HREC

Are you male or female?		
Answer Options	Response Percent	Response Count
Male	39.3%	57
Female	60.7%	88
	<i>answered question</i>	145
	<i>skipped question</i>	0



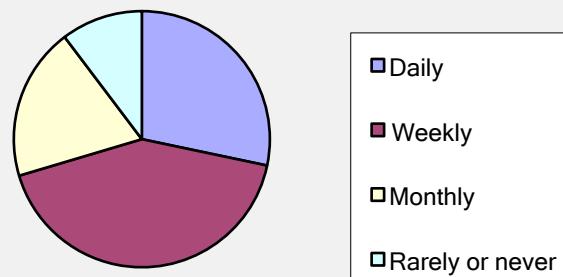
What is your age?		
Answer Options	Response Percent	Response Count
21-25	31.0%	45
26-30	32.4%	47
31-40	25.5%	37
41-50	6.9%	10
51-60	4.1%	6
61 or over	0.0%	0
	<i>answered question</i>	145
	<i>skipped question</i>	0



How often to do you watch music videos online (eg. on Youtube)?

Answer Options	Response Percent	Response Count
Daily	28.3%	41
Weekly	42.1%	61
Monthly	19.3%	28
Rarely or never	10.3%	15
	<i>answered question</i>	145
	<i>skipped question</i>	0

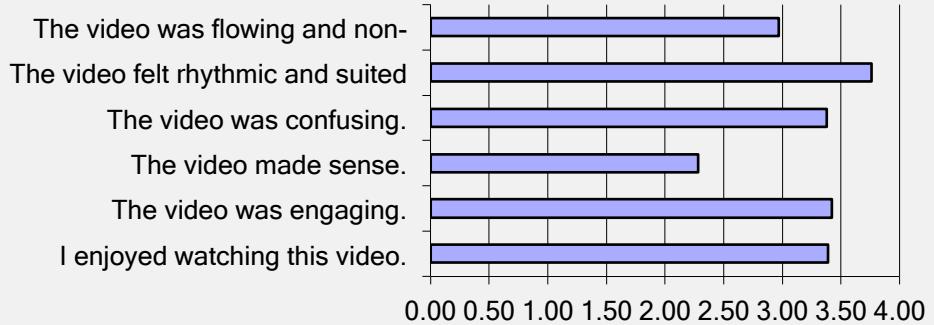
How often to do you watch music videos online
(eg. on Youtube)?



Please rate the following video. video01 from chutopia on Vimeo.

Answer Options	Strongly Disagree	Disagree	Neither agree nor disagree	Agree	Strongly Agree	Rating Average	Response Count
I enjoyed watching this video.	13	6	17	46	11	3.39	93
The video was engaging.	8	12	20	39	14	3.42	93
The video made sense.	23	34	23	13	0	2.28	93
The video was confusing.	2	19	25	36	11	3.38	93
The video felt rhythmic and suited the music.	0	13	14	48	18	3.76	93
The video was flowing and non-disruptive.	10	27	19	30	7	2.97	93
						<i>answered question</i>	93
						<i>skipped question</i>	52

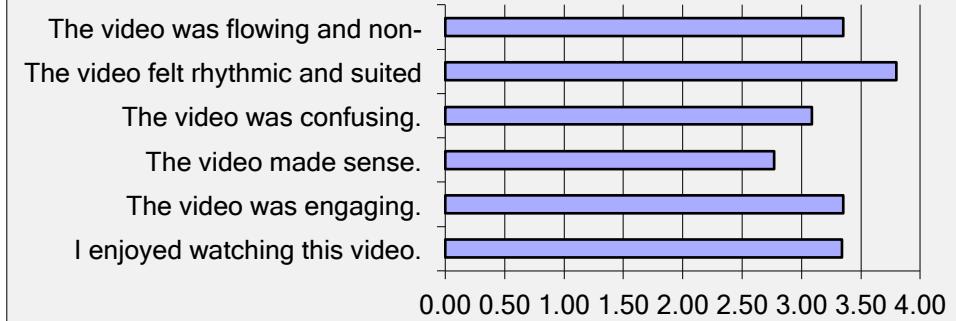
Please rate the following video. video01 from chutopia on Vimeo.



Please rate the following video. video02 from chutopia on Vimeo.

Answer Options	Strongly Disagree	Disagree	Neither agree nor disagree	Agree	Strongly Agree	Rating Average	Response Count
I enjoyed watching this video.	11	9	19	45	9	3.34	93
The video was engaging.	6	13	24	42	8	3.35	93
The video made sense.	13	28	22	27	3	2.77	93
The video was confusing.	2	34	21	26	10	3.09	93
The video felt rhythmic and suited the music.	1	10	12	54	16	3.80	93
The video was flowing and non-disruptive.	3	19	21	42	8	3.35	93
<i>answered question</i>						93	
<i>skipped question</i>						52	

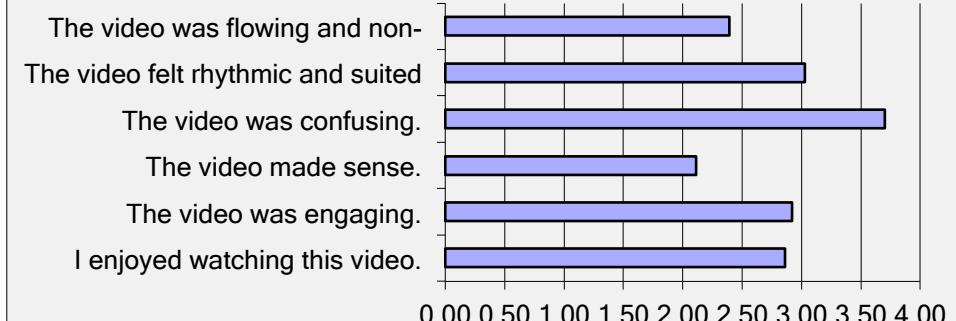
Please rate the following video. video02 from chutopia on Vimeo.



Please rate the following video. video03 from chutopia on Vimeo.

Answer Options	Strongly Disagree	Disagree	Neither agree nor disagree	Agree	Strongly Agree	Rating Average	Response Count
I enjoyed watching this video.	15	23	21	28	6	2.86	93
The video was engaging.	9	25	26	30	3	2.92	93
The video made sense.	22	51	10	8	2	2.11	93
The video was confusing.	1	11	18	48	15	3.70	93
The video felt rhythmic and suited the music.	11	24	16	35	7	3.03	93
The video was flowing and non-disruptive.	22	36	14	19	2	2.39	93
<i>answered question</i>						93	
<i>skipped question</i>						52	

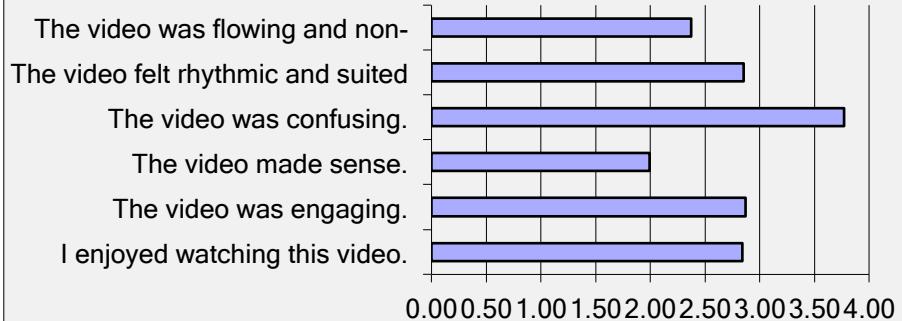
Please rate the following video. video03 from chutopia on Vimeo.



Please rate the following video. video04 from chutopia on Vimeo.

Answer Options	Strongly Disagree	Disagree	Neither agree nor disagree	Agree	Strongly Agree	Rating Average	Response Count
I enjoyed watching this video.	19	20	17	31	6	2.84	93
The video was engaging.	15	25	15	33	5	2.87	93
The video made sense.	27	50	8	6	2	1.99	93
The video was confusing.	1	9	21	41	21	3.77	93
The video felt rhythmic and suited the music.	12	28	21	26	6	2.85	93
The video was flowing and non-disruptive.	19	40	17	15	2	2.37	93
						<i>answered question</i>	93
						<i>skipped question</i>	52

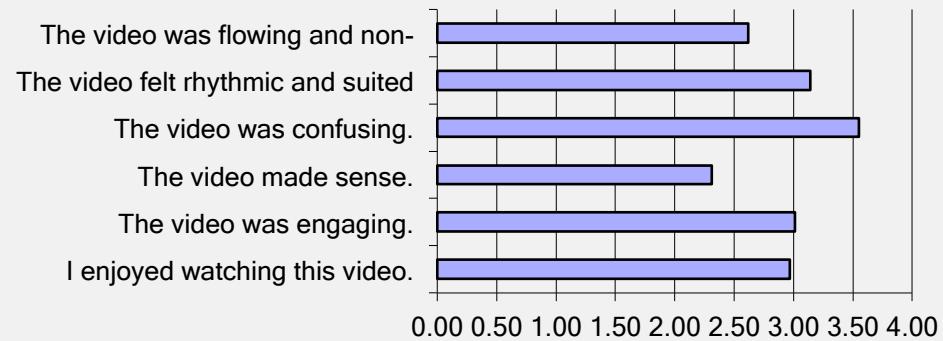
Please rate the following video. video04 from chutopia on Vimeo.



Please rate the following video. video05 from chutopia on Vimeo.

Answer Options	Strongly Disagree	Disagree	Neither agree nor disagree	Agree	Strongly Agree	Rating Average	Response Count
I enjoyed watching this video.	18	12	23	35	5	2.97	93
The video was engaging.	13	21	16	38	5	3.01	93
The video made sense.	20	38	23	10	2	2.31	93
The video was confusing.	2	12	28	35	16	3.55	93
The video felt rhythmic and suited the music.	9	20	20	37	7	3.14	93
The video was flowing and non-disruptive.	15	31	24	20	3	2.62	93
						<i>answered question</i>	93
						<i>skipped question</i>	52

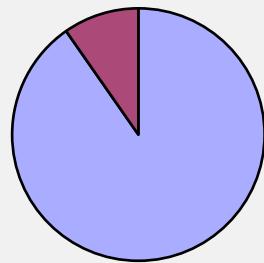
Please rate the following video. video05 from chutopia on Vimeo.



Could you tell a difference between the five videos on this page?

Answer Options	Response Percent	Response Count
Yes	90.3%	84
No	9.7%	9
	<i>answered question</i>	93
	<i>skipped question</i>	52

Could you tell a difference between the five videos on this page?

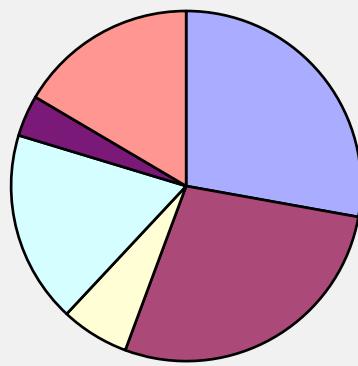


- Yes
- No

Out of the five videos on this page, was there a video you liked in particular? (optional)

Answer Options	Response Percent	Response Count
Video01 (question 4)	27.8%	22
Video02 (question 5)	27.8%	22
Video03 (question 6)	6.3%	5
Video04 (question 7)	17.7%	14
Video05 (question 8)	3.8%	3
Other (please specify)	16.5%	13
	<i>answered question</i>	79
	<i>skipped question</i>	66

Out of the five videos on this page, was there a video you liked in particular? (optional)

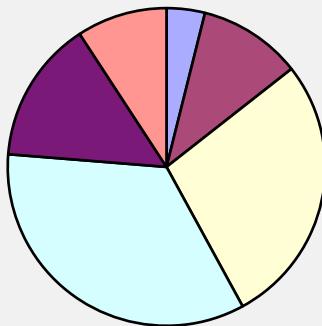


- Video01 (question 4)
- Video02 (question 5)
- Video03 (question 6)
- Video04 (question 7)
- Video05 (question 8)
- Other (please specify)

Out of the five videos on this page, did you dislike one in particular? (optional)

Answer Options	Response Percent	Response Count
Video01 (question 4)	3.9%	3
Video02 (question 5)	10.5%	8
Video03 (question 6)	27.6%	21
Video04 (question 7)	34.2%	26
Video05 (question 8)	14.5%	11
Other (please specify)	9.2%	7
	<i>answered question</i>	76
	<i>skipped question</i>	69

Out of the five videos on this page, did you dislike one in particular? (optional)



- Video01 (question 4)
- Video02 (question 5)
- Video03 (question 6)
- Video04 (question 7)
- Video05 (question 8)
- Other (please specify)

Would you like to add a further comment? (eg. why did you like or dislike particular videos, other thoughts, etc) (optional)

Answer Options	Response Count
	53
<i>answered question</i>	53
<i>skipped question</i>	92

I didn't realise there were so many subplots to the Gangnam Style video!

I prefer to watch less different scenes and relevant scenes.

Personal reasons. Video05 seemed the most disrupted to me.

I just listened to Gangman Style 5 times, I'm not sure if I'll be able to sleep tonight.

The more weird and out of place things, like horses and lone buildings made the video just that little bit more funny and not so annoying

I must have seen this video a million times BEFORE this. The video was way too popular for me not to notice each videos editing. Also, as I don't speak Korean the video was never really going to make sense to me.

Is this study covered under the Geneva Convention? Five HS videos is just mean!

Didn't affect me one way or the other. I wouldn't bother watching the full 3 minute version If there is one and wouldn't buy it.

I disliked number 2 the most because I felt bored and disconnected watching it watching. Number 4 was my favourite perhaps due to the vast variety of scenes, the amusing qualities of these scenes and the long "sumo sauna" scene which I felt was longer than the other videos which also depicted this scene.

Random images that didn't flow was irritating when there weren't many. But when the whole video was only random images it worked.

I disliked video 5 because it was inconsistent with the scenes, too many sporadic images. I liked video 2 because it was telling a story, yet referencing other things which could have logically occurred simultaneously preempting a future scene

1) Did not understand the clip with horse head appearing for no reason or link. 2) Confusing when no link between male character exiting the train, other clip when they did meet up a little better but still did not flow well 3) Dislike the female bums popping up on screen for no apparent reason. 4) funnier sections of the clip the better, little boy, crazy lift guy dancing, car park dance-off, the spa scene would have been funnier if it made sense.

After watching them all in a row, you can't remember all the slight differences between them. However some were more engaging than others. I found the videos that told a story were the best. The ones with just random really quick nonsensical clips after one another were kind of annoying.

Some clips were more in sync with the music, both the actions of the actors and the cutting between scenes. Other videos felt more disconnected from the music such as when a slow motion part of the video was shown during a rapid beat of the music.

I hate pop music and this reflects my attitude when watching this; I can't take this kind of music seriously; I like Death Grips, Scott Walker, Animal Collective - stuff like that

I think it makes no sense the music is very catchy but can't understand girls dancing, guy in yellow suit and on top of someone in elevator

I like it when the 'story' is clear and connected and fits the music.

I wouldn't say I 'liked' or 'disliked' the videos so much as I would say it was clear that the clips had been cut up with varying degrees of correlation between the audio and the visual. It seemed to be that the second video had the strongest correlation. I noticed it most during the 'build-up' section where the screen cut between close-shots of PSY and his leading lady until we saw them dance together. In that version of the video, the visual was timed correctly with the rhythmic interest in the song and resolved at the refrain with a wide shot in the visual.

I really dislike this song so that may have swayed my comments.

Some of them were hard to be engaged in when they felt like they didn't flow or go with the music

They all mostly flowed re vision to the rhythm of the music but some had totally incongruous vision, and some had vision cuts that didn't relate to the beat or the essence of the visual 'story' script. Good luck - great research

I didn't understand the scene in the elevator

base on the music video background and the theme of the screen

K-pop isn't my kind of style. Despite that I am Asian, I find this ridiculous

what was I supposed to be looking for? I just enjoyed each for what it was.

I disliked the one that kept changing the shots and didn't have a natural flow to it. It was distracting and didn't make sense to me - whereas the one I did like had more of a storyline to it and was more enjoyable to watch because of this. I felt like I could connect with the video more, opposed to feeling detached and confused.

02 was good because the sequence was good. It was very obvious in 03 that the sequences and flow of action and interaction were out of order / disjointed

I did not understand or like the water scenes that were included in some of the videos. The girl scenes were all good but could have been longer. The dance off was a good opening and together with the girl scenes videos 1 and 2 were my favourite.

it wasn't in sync and I disliked all the videos with close ups on girls bottoms and slutty dancing

Some of the scenes didn't seem to match the music, and some did feel 'patchier' than others.

By the time you get to the third one it's boring and made no sense

A couple of internal observations: I was more inclined to enjoy the earlier ones than the later ones because I hadn't just listened to the same minute of music five times over. Also, you don't expect a pop music video to make sense - hot babes and attitude are way more valuable than narrative cohesion.

#2 was too choppy with slow mo, it makes me feel motion sick LOL

Video05(Q8) seem really disjointed - video footage didn't seem to match the rhythm of the music hardly ever. Video2(Q5) seemed the closest match between video and music.

I don't feel like the term of sense really applies to any of the videos displayed as, I cannot understand the lyrics and I feel I would derive the feeling of sense making from the words in relation to the images.

There seemed to be more story in the second version, I don't necessarily need story to enjoy music video but when they are using so many images from various locations I much prefer some sort of continuity otherwise I lose interest very quickly

Other than the order of the scenes, the videos were quite similar. The fact that it is not in English, or any of the other languages I understand, did contribute towards its confusing message.

Most people just love the fun and the catchy hook. The song is not in English, so for most English speaking people the video doesn't need to represent any serious message. The video is fun and non-sensical like the song.

Disliked it when the visual scene was not edited to the beat, on the beat precisely. Also disliked it when there were too many themes and ideas introduced without elaboration or any continuation of the theme/idea. Enjoyed and liked having some random shots and ideas/themes except too many became haphazard - otherwise there wasn't enough variation...

I probably didn't like the ones that kept jumping to completely different scenes with no links between scenes at all
felt all over the place

I think it was #3.. disliked it because it seemed a little more choppy/changey than the others.. shorter sequences/quick shots jumping from one thing to another. Hope this helps :)

Probably would have helped if I hadn't watched this video a million times before doing this survey though!

I hate that song with a passion...why did you put me through that

In diesem Fall beeinflusst die unterschiedliche Zusammenstellung der Szenen meine Einstellung zum Video nicht. Die verschiedenen Videos evozieren - trotz unterschiedlichem Ablauf - dieselben Gefühle.

it is definitely hard to pin down the differences after watching more than two versions. after watching all of them, I couldn't remember the specific differences anymore

My observation, the fifth video may have been cut to show the chaotic dance moves/jigging closer together... whilst still kind of puzzling (in a fun way), it at least had that consistency vs. scenes with walking and posing.

I really didn't like video 2. There were too many slow motion parts for this rhythmic music. In video 1 and 3, I think the very short scene where you can see the horse is a bit confusing. It felt as it was just pasted in without any reason.

Main reason for liking or disliking was choreography/rhythmic synchronisation.

As the whole Gangam style video generally makes no sense at all, it might be a bad choice for this survey. Apart from video02, all other videos jump completely random between unlinked scenes; you can't follow any storyline/s could'n get the fifth to play.

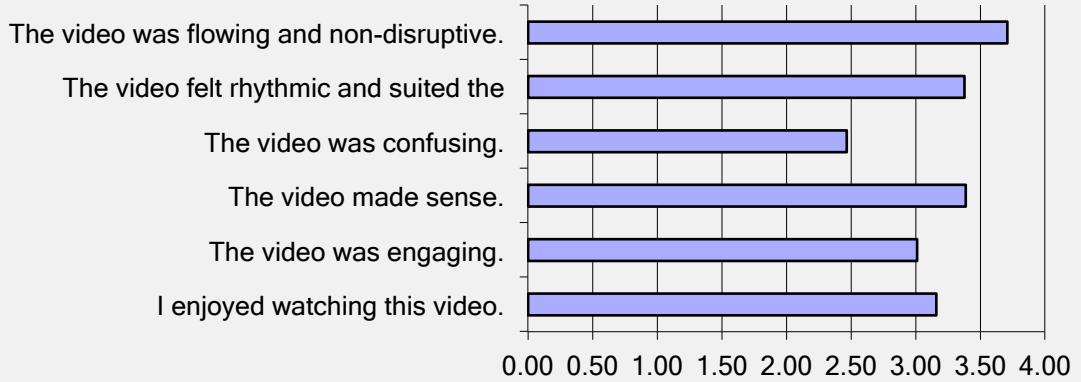
The clips seem to have been taken out of context, the editing seemed rushed and nonsensical

I wanted see more of that horse in the first video, to be honest.

Please rate the following video. video06 from chutopia on Vimeo.

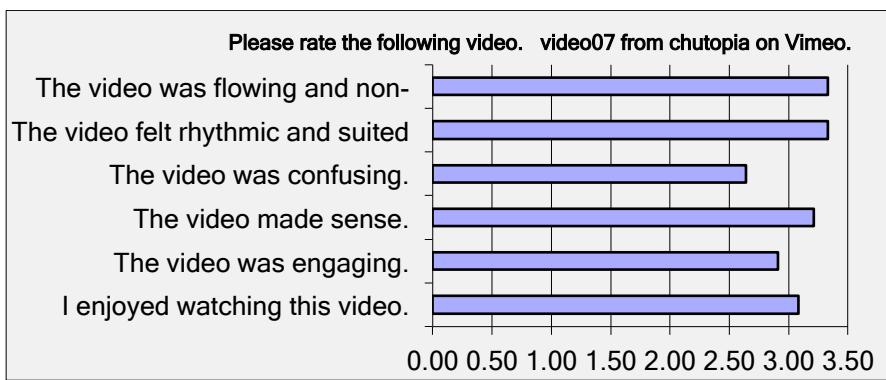
Answer Options	Strongly Disagree	Disagree	Neither agree nor disagree	Agree	Strongly Agree	Rating Average	Response Count
I enjoyed watching this video.	4	22	16	26	8	3.16	76
The video was engaging.	7	19	21	24	5	3.01	76
The video made sense.	6	9	19	33	9	3.39	76
The video was confusing.	11	31	23	9	2	2.47	76
The video felt rhythmic and suited the music.	5	12	16	35	8	3.38	76
The video was flowing and non-disruptive.	1	6	16	44	9	3.71	76
answered question						76	
skipped question						69	

Please rate the following video. video06 from chutopia on Vimeo.



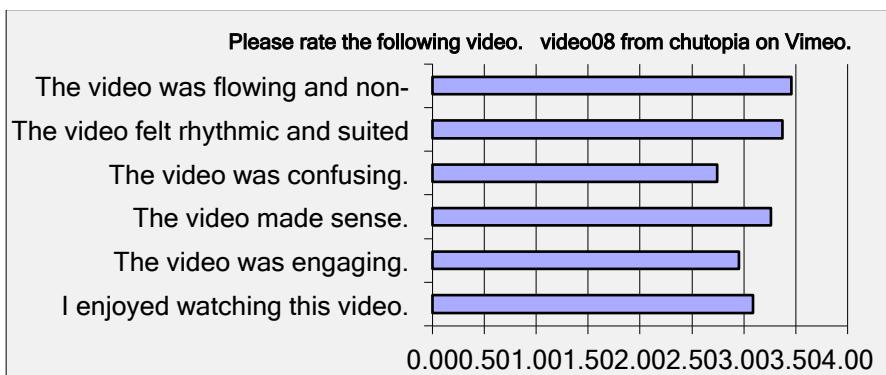
Please rate the following video. video07 from chutopia on Vimeo.

Answer Options	Strongly Disagree	Disagree	Neither agree nor disagree	Agree	Strongly Agree	Rating Average	Response Count
I enjoyed watching this video.	4	16	27	28	1	3.08	76
The video was engaging.	6	23	19	28	0	2.91	76
The video made sense.	2	16	23	34	1	3.21	76
The video was confusing.	4	36	19	17	0	2.64	76
The video felt rhythmic and suited the music.	3	12	23	33	5	3.33	76
The video was flowing and non-disruptive.	1	17	17	38	3	3.33	76
<i>answered question</i>						76	
<i>skipped question</i>						69	



Please rate the following video. video08 from chutopia on Vimeo.

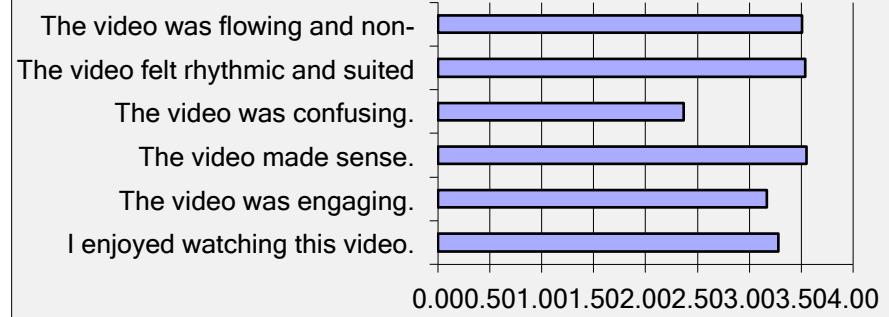
Answer Options	Strongly Disagree	Disagree	Neither agree nor disagree	Agree	Strongly Agree	Rating Average	Response Count
I enjoyed watching this video.	4	19	24	24	5	3.09	76
The video was engaging.	5	23	21	25	2	2.95	76
The video made sense.	4	14	19	36	3	3.26	76
The video was confusing.	3	38	15	16	4	2.74	76
The video felt rhythmic and suited the music.	3	14	17	36	6	3.37	76
The video was flowing and non-disruptive.	2	12	17	39	6	3.46	76
<i>answered question</i>						76	
<i>skipped question</i>						69	



Please rate the following video. video09 from chutopia on Vimeo.

Answer Options	Strongly Disagree	Disagree	Neither agree nor disagree	Agree	Strongly Agree	Rating Average	Response Count
I enjoyed watching this video.	4	16	17	33	6	3.28	76
The video was engaging.	5	14	23	31	3	3.17	76
The video made sense.	2	8	19	40	7	3.55	76
The video was confusing.	7	46	13	8	2	2.37	76
The video felt rhythmic and suited the music.	3	11	10	46	6	3.54	76
The video was flowing and non-disruptive.	2	7	21	42	4	3.51	76
						answered question	76
						skipped question	69

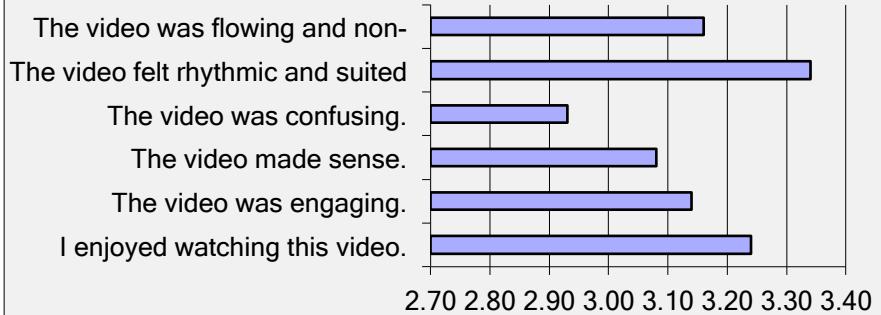
Please rate the following video. video09 from chutopia on Vimeo.



Please rate the following video. video10 from chutopia on Vimeo.

Answer Options	Strongly Disagree	Disagree	Neither agree nor disagree	Agree	Strongly Agree	Rating Average	Response Count
I enjoyed watching this video.	6	14	19	30	7	3.24	76
The video was engaging.	7	16	16	33	4	3.14	76
The video made sense.	4	19	24	25	4	3.08	76
The video was confusing.	5	25	18	26	2	2.93	76
The video felt rhythmic and suited the music.	4	13	18	35	6	3.34	76
The video was flowing and non-disruptive.	5	21	12	33	5	3.16	76
						answered question	76
						skipped question	69

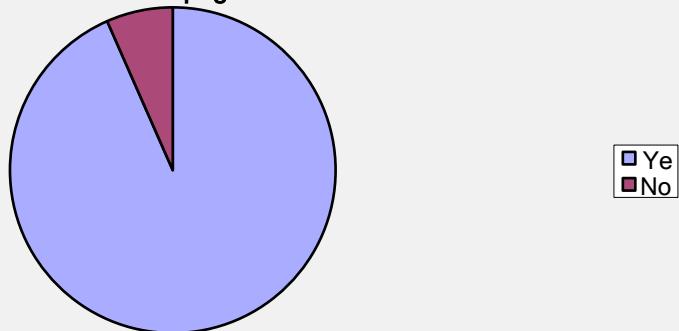
Please rate the following video. video10 from chutopia on Vimeo.



Could you tell a difference between the five videos on this page?

Answer Options	Response Percent	Response Count
Yes	93.4%	71
No	6.6%	5
	<i>answered question</i>	76
	<i>skipped question</i>	69

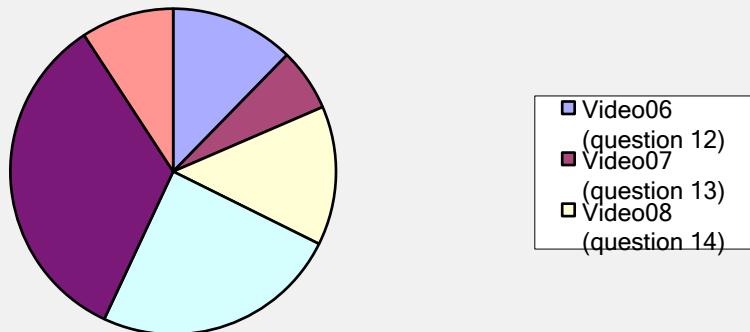
Could you tell a difference between the five videos on this page?



Out of the five videos on this page, was there one you liked in particular? (optional)

Answer Options	Response Percent	Response Count
Video06 (question 12)	12.3%	8
Video07 (question 13)	6.2%	4
Video08 (question 14)	13.8%	9
Video09 (question 15)	24.6%	16
Video10 (question 16)	33.8%	22
Other (please specify)	9.2%	6
	<i>answered question</i>	65
	<i>skipped question</i>	80

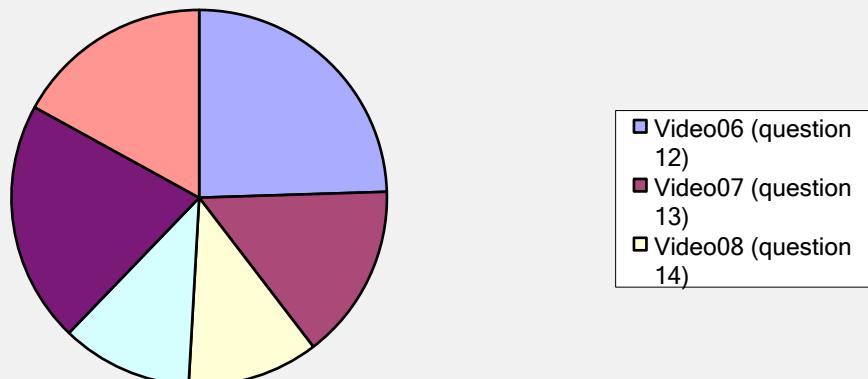
Out of the five videos on this page, was there one you liked in particular? (optional)



Out of the five videos on this page, did you dislike one in particular? (optional)

Answer Options	Response Percent	Response Count
Video06 (question 12)	24.5%	13
Video07 (question 13)	15.1%	8
Video08 (question 14)	11.3%	6
Video09 (question 15)	11.3%	6
Video10 (question 16)	20.8%	11
Other (please specify)	17.0%	9
	<i>answered question</i>	53
	<i>skipped question</i>	92

**Out of the five videos on this page, did you dislike one in particular?
(optional)**



**Would you like to add a further comment? (eg. why did you like or dislike particular videos, other thoughts, etc)
(optional)**

Answer Options	Response Count
	34
<i>answered question</i>	34
<i>skipped question</i>	111

One of the videos didn't work properly. It stopped after about 15 seconds and went to the end.

I like this style!

I liked the combination of light and the different scenes in Video10.

I felt 10 best communicated the feeling/message of the song. It made me feel feels.

The video story was what the song could be about. Very youthful and spontaneous scenes of the character doing things.

I did not particularly like any of the videos, mainly because of lack of flow/story.

The song affects how I feel about the video. I don't like this song.

I liked the last one because of the colourful candy scene.

The slow images were lovely, especially in the moving car.

I liked video 10 because it was a little bit more chaotic; challenged my mind more to figure out all the clues. Video 06 was boring and drawn out, just a sad girl at work caught up in the monotony.

1) Not seeing the whole video the viewer is left without a climax 2) It is boring, sorry. 3) I could see parts of the clip were attempting to connect travelling to a destination, however was that the car ride only? 4) Enjoyed Q16 attempted to link the male character 5) Enjoyed parts of the melanchololy, just feel like it needed some form of either happiness or show why there is sadness. Love the song Though!

I felt these five videos had a strong narrative element to them that really suited the music.

They were all cut differently but got the message across of the daydreaming girl yearning to be taken away for a better life

I can't understand what a \$2 store has to do with a jet plane

Sorry, No. 7 wouldn't load. This song is less interesting when played with the music video. It's a slow song, but the video makes it feel like it's dragging.

I like it when scenes last comparative to the tone of the music. It is distracting to watch quick transitions with mellow music. Equally, it can be a little boring when scenes drag on too long...

The first video was actually boring and vision seemed disconnected from scene to scene. Video 10 was the most connected and it made a difference seeing the guy singing

boring screen of the music background

all of the videos were quite dull and depressing, but I liked the use of color, and I really enjoyed the videos that contained the parallel to her being sad to her dreaming of happiness.

06 not enough contrast in tempo - too much time spent on boring bits - too solemn 07 nice opening - more interesting combination of shots.

I did not like the videos that did not have context - a direct connection between the girls wanting to travel and the song. Otherwise the shots of the store made little sense.

pretty boring, just focuses on a girl moping around a store.

Some were very engaging and had me wanting to know more about her story, some didn't hand the idea of her wanting to escape as easily. I thought Vid 9 had a good balance and I enjoyed it.

The first video you watch with no frame of reference, so you answer according to a scale you invent in your own head. Then any subsequent videos have to be measured relative to the first one, and I think this can compromise the results. It'd be interesting to see if people's responses would differ if you had them watch three or four sample videos first to demonstrate what they'd be responding to, and THEN showed them the actual examples you were testing.

I particularly disliked number one because it was so slow I couldn't wait for it to end. I liked number 9 the most but there were some parts that held onto the scene a little too long for me to like it more. If the scenes weren't so boring it probably wouldn't bother me so much. Some scenes there was more happening I didn't mind those ones, but the scenes with not much happening just felt dragged out.

I didn't like the way 3 of the videos cut off the beat or even lyrics, they seemed to just drag on making no sense. The 2 that cut on beat and lyric flowed much better and were much more enjoyable to watch. I don't know if it helps but the girl in the video has the most beautiful eyes.

I felt that video 9 had a straight story line whereas Video 10 jumped between scenes that didn't seem to show a story.

I think they could have related the monotony of her life and her urge to escape a bit more strongly with the imagery

i liked video 6 the most, however video 10 was close behind, however i liked how that one had video of the artists singing

This video (and song) is relatively mundane and unenjoyable in each version

Im Gegensatz zum ersten Video besteht hier ein Zusammenhang zwischen der Zusammenstellung der Szenen und den Bewertungen des Stimulusmaterials.

In my view, only the first successfully captured what I believe to be the theme... escapism/day-dreaming in the setting of a monotonous retail job.

I really liked video 10. Not only because it was way different to the others but also that one can feel that the girl wants to go away and be free (at least for me).

Video09 is the only one that makes any sense the others just show some pictures in random order

Do you have any other comments?

Answer Options	Response Count
	18
<i>answered question</i>	18
<i>skipped question</i>	127

I prefer to watch the second kind of style of mv~

Cheese is a kind of meat. A healthy yellow beef.

That was fun thanks :)

Interesting survey; I would find it interesting to learn what you were actually looking at.

No

not really

All the best with the development of the program - it will be interesting to see videos made with it.

Good luck!

no

Random images in music video have always been a part of the art form but it is really nice to be able to find some reason for them

It was difficult to differentiate between the videos on one quick viewing - hope I helped your research

The Gangam Style vid is obviously ultra-famous for being out there and wacky... which I can get on board with. The second video I haven't seen/heard before... but this survey will compel me to check it out since I think it's cool.

Good luck for your PhD :)

As a musician I should state that I have a background in music.

Interesting topic

Nothing other than that I'm already sick of Gangnam Style.

Appendix C

Version three source code

```
createSequence.php

<?php

    // Create Database Connection

    /*
        DATABASE ACCESS FOR TABLE "division02"
        WITH $connection
    */

    //save all clips in an array
    $clipQuery = "SELECT * FROM division02";
    $tempClips = mysqli_query($connection, $clipQuery);

    if (!$tempClips) die("Database query failed.");

    $rawClips = array();
    $clips = array();
    $segmentClipIndexes = array(array());

    while ($row = mysqli_fetch_assoc($tempClips)) $rawClips[] = $row;

    function array_push_assoc($array, $key, $value) {
        $array[$key] = $value;
        return $array;
    }

    $songDuration = 217.064583;
    $segmentInterval = 10;
    $minimumDuration = 0.30;

    $userInput = false;

    if (isset($_POST['segmentInput'])) {
```

```

if($_POST['segmentInput'] != (-1)) {
    $segmentInterval = $_POST['segmentInput'];
    $userInput = true;
}

// add fields to existing database information
foreach($rawClips as $k => $rawClip) {
    $tempClip = array_push_assoc($rawClip, "hasPlayed", 0);
    $segment = floor($rawClip["startTime"] / $segmentInterval);
    $segmentClipIndexes[$segment][] = $k;
    $clips[] = $tempClip;
}

//get probability values
$probabilityQuery = "SELECT * FROM probability";
$tempProb = mysqli_query($connection, $probabilityQuery);
$probabilityValues = array();
if(!$tempProb) die("Database query failed at probability.");
while($probRow = mysqli_fetch_assoc($tempProb)) $probabilityValues[] = $probRow;

$probEmily;
$probSep;
$probOtherBand;

$probEmilyCurrent;
$probSepCurrent;
$probOtherBandCurrent;

$songIteration = 0;
$performanceOnly = "off";
$emilyOnly = "off";

planSequence();

// -----
function getProbability($interval, $probType) {

```

```

global $songDuration;
global $segmentInterval;
global $probabilityValues;

$prob = array();

for($j = 0; $j < $songDuration; $j += $segmentInterval) {

    $yy = 0;
    $yn = 0;
    $ny = 0;
    $nn = 0;

    foreach($probabilityValues as $k => $probVal) {

        if($probVal["timecode"] >= $j && $probVal["timecode"] < ($j + $segmentInterval)) {

            switch($probType) {
                case 0:
                    switch($probVal["emily"]){
                        case 0:
                            $yy++;
                            break;
                        case 1:
                            $yn++;
                            break;
                        case 2:
                            $ny++;
                            break;
                        case 3:
                            $nn++;
                            break;
                    }
                break;
                case 1:
                    switch($probVal["sep"]){
                        case 0:
                            $yy++;
                            break;
                        case 1:
                            $yn++;
                            break;
                    }
                break;
            }
        }
    }
}

```

```

break;
case 2:
$ny++;
break;
case 3:
$nn++;
break;

}
break;
case 2:

switch($probVal["otherBand"]){
case 0:
$yy++;
break;
case 1:
$yn++;
break;
case 2:
$ny++;
break;
case 3:
$nn++;
break;

}
break;
}

$prob[] = array(array($nn, ($nn+$ny) , array($yn, ($yn+$yy))));

}

return $prob;
}

// -----
function planSequence(){
global $probEmily;
global $probSep;
}

```

```

global $probOtherBand;
global $probEmilyCurrent;
global $probSepCurrent;
global $probOtherBandCurrent;
global $songIteration;
global $songDuration;
global $segmentInterval;
global $clips;
global $segmentClipIndexes;
global $minimumDuration;
global $performanceOnly;
global $emilyOnly;
$plannedClip = (-1);
$previousClip = (-1);
$performanceOnly = (-1);
$segmentShots = array();
$segment = (-1);
$probChoice = array(50, 100);

if(isset($_POST['orderInput'])){
    if($_POST['orderInput'] != (-1)){
        $order = $_POST['orderInput'];
        $random = 100 - $order;
        $probChoice = array($order, $random);
    }
}

$editTimes = array();
$shotDurations = array();
$finalSequence = array();
$clipStarts = array();

if(isset($_POST['performanceOnly'])){

```

```

$performanceOnly = $_POST['performanceOnly'];
$userInput = true;
}

if(isset($_POST['emilyOnly'])){
    $emilyOnly = $_POST['emilyOnly'];
    $userInput = true;
}

if($userInput){

    if(isset($_POST['oldclips']) && isset($_POST['oldClipStarts']) && isset($_POST['oldDurations']) && isset($_POST['oldEditTimes'])){

        foreach($_POST['oldclips'] as $k => $oldClip) $finalSequence[] = $oldClip;

        foreach($_POST['oldClipStarts'] as $k => $oldClipStart) $clipStarts[] = $oldClipStart;

        foreach($_POST['oldDurations'] as $k => $oldDuration) $durations[] = $oldDuration;

        foreach($_POST['oldEditTimes'] as $k => $oldEditTime) $editTimes[] = $oldEditTime;

        $songIteration = end($_POST['oldEditTimes']) + end($_POST['oldDurations']);

        $plannedClip = end($_POST['oldClips']);
        $plannedDuration = end($_POST['oldDurations']);
    }

    foreach($finalSequence as $k => $sequenceClip){

        if(!$_clips[$sequenceClip]["hasPlayed"]) $clips[$sequenceClip]["hasPlayed"] = 1;

    }
}

if($performanceOnly == "on"){

    foreach($segmentClipIndexes as $k => $segments){

        $probEmily[] = array(array(0, 0), array(0, 0));
        $probSep[] = array(array(1, 2), array(1, 2));
        $probOtherBand[] = array(array(1, 2), array(1, 2));

    }
}

elseif($emilyOnly == "on"){
}
}

```

```

foreach($segmentClipIndexes as $k => $segments) {
    $probEmily[] = array(array(1, 2), array(1, 2));
    $probSep[] = array(array(0, 0), array(0, 0));
    $probOtherBand[] = array(array(0, 0), array(0, 0));
}

} else{
    $probEmily = getProbability($segmentInterval, 0);
    $probSep = getProbability($segmentInterval, 1);
    $probOtherBand = getProbability($segmentInterval, 2);
}

do{
    // 1 -- assign probability values according to musicTime
    foreach($segmentClipIndexes as $k => $clipIndex){
        $timeframe = $k * $segmentInterval;

        if($songIteration >= $timeframe && $songIteration < ($timeframe + $segmentInterval)) {

            $segment = $k;
            break;
        }
    }

    $probEmilyCurrent = $probEmily[$segment];
    $probSepCurrent = $probSep[$segment];
    $probOtherBandCurrent = $probOtherBand[$segment];

    $choiceNumber = rand(1, end($probChoice));
    $choiceType = (-1);

    if($choiceNumber <= $probChoice[0]) {
        $choiceType = "order";
    } elseif($choiceNumber > $probChoice[0] && $choiceNumber <= $probChoice[1]){
}
}

```

```

$choiceType = "random";

}

// 2.1 -- set first clip if musicTime=0
if($songIteration){

    $plannedClip = clipDecision($previousClip, $segment, $finalSequence, $choiceType, $editTimes, $shotDurations, $songIteration,
        $performanceOnly);

    // 2.2 -- set all other clips if musicTime>0
} elseif($songIteration > 0){

    $previousClip = $plannedClip;
    $plannedClip = clipDecision($previousClip, $segment, $finalSequence, $choiceType, $editTimes, $shotDurations, $songIteration);

}

// 5 -- set sequence
$finalSequence[] = $plannedClip;

// 5.1 -- if clip performance

if($clips[$plannedClip]["sep"] == '1' || $clips[$plannedClip]["otherBand"] == '1'){

    if($songIteration > 0){

        $newPreviousDuration = $clips[$plannedClip]["startTime"] - end($editTimes);

        if($newPreviousDuration >= $minimumDuration){

            array_pop($shotDurations);
            $shotDurations[] = $newPreviousDuration;

            $songIteration = $clips[$plannedClip]["startTime"];
            $editTimes[] = $clips[$plannedClip]["startTime"];
            $clipStarts[] = 0;
            $shotDurations[] = $clips[$plannedClip]["duration"];

        }elseif(((($newPreviousDuration + $clips[$plannedClip]["duration"]) / 2) >= $minimumDuration){

            array_pop($shotDurations);
            $shotDurations[] = $minimumDuration;

        }

    }

}

}

```

```

$songIteration = end($editTimes) + end($shotDurations);
$editTimes[] = end($editTimes) + end($shotDurations);
$clipStarts[] = end($editTimes) - $clips[$plannedClip]["startTime"];
$shotDurations[] = $clips[$plannedClip]["duration"] - end($clipStarts);
}

} else{

    $editTimes[] = 0;
    $clipStarts[] = 0;
    $shotDurations[] = $clips[$plannedClip]["duration"];
}

} else{

    $editTimes[] = $songIteration;
    $clipStarts[] = 0;
    $shotDurations[] = $clips[$plannedClip]["duration"];
}

// 6 -- set current clip to played
$clips[$plannedClip]["hasPlayed"] = 1;

// 7 -- set previous
$previousClip = $plannedClip;

// 9 -- progress musicTime and counter ($)
$songIteration += end($shotDurations);

// 10 -- escape loop if too long
if(count($finalSequence) > 250) break;

}while($songIteration <= $songDuration);

$sequenceData = array("clips" => $finalSequence,
                      "starts" => $editTimes,
                      "durations" => $shotDurations,
                      "clipStarts" => $clipStarts);

echo json_encode($sequenceData);

}

```

```

// ----- function clipDecision($previousClip, $segment, $finalSequence, $choiceType, $editTimes, $shotDurations, $currentTime) {
global $clips;

$set1 = makeSet(1, $previousClip, $segment, $finalSequence, $editTimes, $shotDurations);

$playNext = (-1);

if($choiceType == "random"){

    $choice = rand(0, (count($set1) - 1));

    $playNext = $set1[$choice];

} elseif($choiceType == "order"){

    if($previousClip == (-1)) {

        $playNext = $set1[0];

    } else{

        $tempArray = sortSequence($clips, $set1, "startTime", SORT_ASC);

        $playNext = getClosest($currentTime, $tempArray);

    }

}

return $playNext;

}

// ----- function getClosest($search, $array) {
global $clips;

$closest = null;

foreach($array as $k => $v){

    // tests if the difference between the previous clip and search is bigger than current val and search
    if($closest == null || abs($search - $clips[$v]["startTime"]) > abs($clips[$closest]["startTime"] - $search)){

        $closest = $v;

    }

}

```

```

        }
    }

    return $closest;
}

// -----
function sortSequence($array, $segment, $on, $order= SORT_ASC) {

    $newArray = array();
    $sortableArray = array();

    if(count($array) > 0 && count($segment) > 0) {

        foreach($segment as $key => $clipIndex) {

            $sortableArray[] = $array[$clipIndex][$on];
        }

        switch ($order) {
            case SORT_ASC:
                asort($sortableArray);
                break;
            case SORT_DESC:
                arsort($sortableArray);
                break;
        }

        foreach ($sortableArray as $k => $v) {
            $newArray[$k] = $segment[$k];
        }
    }

    return $newArray;
}

// -----
function makeSet($setType, $previousClip, $segment, $finalSequence, $editTimes, $shotDurations) {

```

```

global $clips;
global $probEmilyCurrent;
global $probSepCurrent;
global $probOtherBandCurrent;

switch($setType) {
    case 1:
        $currentParamSet = filterParams($currentParamSet, $previousClip["emily"],
            $clips[$previousClip["sep"]], $clips[$previousClip['otherBand']]);
        $currentProbSetFull = array($probEmilyCurrent, $probSepCurrent, $probOtherBandCurrent);
        break;
    }
}

$clipSet = filterParams($currentParamSet, $currentProbSetFull, $setType, $previousClip, $segment, $finalSequence, $editTimes,
    $shotDurations);

return $clipSet;
}

// -----
function filterParams($paramSet, $probSetFull, $setType, $prevClip, $segment, $finalSequence, $editTimes, $shotDurations) {

    global $clips;
    global $segmentClipIndexes;
    global $songIteration;
    global $performanceOnly;
    global $emilyOnly;

    $filteredShots = array();
    if($prevClip == (-1)) {
        foreach($segmentClipIndexes[$segment] as $k => $v) {
            if($performanceOnly == "on") {
                if($clips[$v]["sep"] == "1" || $clips[$v]["otherBand"] == "1") {
                    if($clips[$v]["startTime"] == 0) $filteredShots[] = $v;
                }
            }
        }
    }
}

```

```

}else if($clips[$v] ["sep"] == '0' && $clips[$v] ["otherBand"] == '0'){

$filteredShots[] = $v;

}

}elseif{

// 1 -- get right prob sets
foreach($paramSet as $k => $param) $probSetPar[] = filterProbSet($param, $probSetFull[$k]);

// 2 -- get next parameters
foreach($probSetPar as $k => $probSet) $nextParam[] = getNextParam($nextParam, $probSet);

// 3 -- transform set, if necessary (according to restriction type)
if(count($nextParam) == 3){

$checkSum = 0;

foreach($nextParam as $k => $v){

// if value = (1), iterate on checkSum
if($v != 0){

$checkSum++;
break;

}

}

$probYesPct = array();

// if checkSum is 0, do some stuff
if($checkSum == 0){

// probSetPar = (thisProbEmily, thisProbSep, thisProbOther)
foreach($probSetPar as $k => $probSet){

$probSum = $probSet[1];
$tempPct = 0;

// if there is a probability for this parameter value, get the probability for yes and normalise to %
if($probSum > 0){

```

```

$yesVal = $probSet[1] - $probSet[0];
$tempPct = 100 / $probSum * $yesVal;
$probYesPct[] = $tempPct;

}else{
    $probYesPct[] = 0;
}

}

// save index for highest value
// maximum presents as an array, but should only ever have one value in it, this is the index of the highest value
$maximum = array_keys($probYesPct, max($probYesPct));

$nextParam[$maximum[0]] = 1;

}

// 4 -- create a shot collection according to above parameters
$filteredShots = collectShots($setType, $nextParam, $prevClip, $segment, $finalSequence, $editTimes, $shotDurations);

}

return $filteredShots;
}

// -----
function collectShots($setType, $nextParam, $previousClip, $segment, $finalSequence, $editTimes, $shotDurations){

global $clips;
global $segmentClipIndexes;
global $minimumDuration;

$filteredShots = array();

foreach($segmentClipIndexes[$segment] as $k => $clipIndex){

    // check if the clip is active
    if($clips[$clipIndex]["extra"] == 0 && $clips[$clipIndex]["inactive"] == '0' && $clips[$clipIndex]["hasPlayed"] != 1 && $clipIndex != $previousClip){

        switch($setType){


```

```

case 1:
$checkParam = array($clips[$clipIndex]["emily"], $clips[$clipIndex]["sep"], $clips[$clipIndex]["otherBand"]);
break;
}

$paramCheckSum = 0;

foreach($checkParam as $k2 => $param) {
if ($nextParam[$k2] == $param) $paramCheckSum++;

if ($paramCheckSum == count($checkParam)) {

if ($clips[$clipIndex]["sep"] == '1' || $clips[$clipIndex]["otherBand"] == '1') {

$currentStart = $clips[$clipIndex]["startTime"];
$currentDuration = $clips[$clipIndex]["duration"];

if ($currentStart > end($editTimes) && $currentStart <= (end($editTimes) + end($shotDurations))) {

$newPreviousDuration = $currentStart - end($editTimes);
if (($newPreviousDuration + $currentDuration) / 2) >= $minimumDuration) $filteredShots[] = $clipIndex;

}
}

}elseif{

$filteredShots[] = $clipIndex;

}
}
}

// if there's nothing in the filtered group, get the whole segment
if (!$filteredShots) {
foreach($segmentClipIndexes[$segment] as $k => $clipIndex) {

```

```

if($clips[$clipIndex]["extra"] == 0 && $clips[$clipIndex]["inactive"] == '0' && $clips[$clipIndex]["hasPlayed"] != 1 && $clipIndex != $previousClip) {
    $passOn = false;

    foreach($finalSequence as $k2 => $finalClipIndex) {

        $currentStart = $clips[$clipIndex]["startTime"];
        $currentDuration = $clips[$clipIndex]["duration"];

        if($clips[$clipIndex]["sep"] == '1' || $clips[$clipIndex]["otherBand"] == '1') {
            if($currentStart > end($editTimes) && $currentStart <= (end($editTimes) + end($shotDurations)) && $currentStart != end($editTimes)) {

                $newPreviousDuration = $currentStart - end($editTimes);
                if((($newPreviousDuration + $currentDuration) / 2) >= $minimumDuration) $passOn = true;
            }
        } else{
            $passOn = true;
        }
    }

    if($passOn) $filteredShots[] = $clipIndex;
}

if(!$filteredShots) {
    foreach($clips as $k => $v) {
        if($v["extra"] == 1 && $v["inactive"] == '0' && $v["hasPlayed"] != 1) $filteredShots[] = $k;
    }
}

```

```

        return $filteredShots;
    }

    // -----
    function getNextParam($currentParamSet, $probSetPar) {
        $probMax = end($probSetPar);
        $restricted = false;
        $nextParam = (-1);

        if (!$restricted) {
            if ($probMax) {
                $nextParam = 0;
            } elseif ($probMax) {
                // this is a point that keeps changing and confusing
                // the current parameter structure is (nn,nn+ny),(yn,yn+yy)
                // N, Y - 0, 1
                $probNumber = mt_rand(1, $probMax);
            }
            foreach($probSetPar as $k => $probVal) {
                //check for i=0 because condition test changes
                if ($k == 0 && $probNumber <= $probVal) {
                    $nextParam = $k;
                } elseif ($k != 0 && $probNumber > $probSetPar[$k-1] && $probNumber <= $probVal) {
                    $nextParam = $k;
                }
            }
        }
        return $nextParam;
    }
}

```

```
// -----  
// function filterProbSet($paramType , $probSet){  
//  
//    for($i = 0; $i < 2; $i++){  
//  
//        if($paramType == $i){ // this is not right  
//  
//            for($j = 0; $j < 2; $j++){  
//  
//                $probSetPar[] = $probSet[$i][$j];  
//  
//            }  
//  
//            break;  
//        }  
//    }  
//  
//    return $probSetPar;  
//}  
?  
>
```

APPENDIX D

Ethics Griffith Ref No. QCA/26/14/HREC

Participant 1

Gender: male

Age group: 31-40

Experience: 4-8

Qu01: They had consistent narrative arcs and used the same style of jump cutting back and forth in time across the different acts.

Qu02: It was consistent with current conventions in music video cutting ie. fast cuts around a loose narrative intercut with performance.

Qu03: No particular parts jumped out at me.

Qu04: For the style of music I think it was appropriate.

Qu05: I think most people would say it was a conventional performance rock clip.

Participant 2

Gender: male

Age group: 41-50

Experience: 16-30

Qu01: I was sick of that song about a minute into the first watch. I'd completely lost feeling in my ears by the third. Really, it sucked very, very hard.

The video content was so unmemorable that it was hard to compare the three versions. They were all variations on a theme of a bland montage of nothing much happening in someone's dingy share house, featuring a bored looking alcoholic woman and some indy-rock dudes who all needed to wash their hair.

Qu02: It looked like the kind of editing you see depressingly often on music videos. It worked rhythmically, but was completely meaningless.

Occasionally there'd be a dodgy cut where you get a shot with a bit of out of focus camera shake and nothing else for 12 frames or so, but overall it looked like a human had cut it. A seventeen year old human.

There seemed to be little difference between the three cuts, style wise.

Qu03: No. My eyes just glazed over at about thirty seconds in and from then it was all just colour and movement.

Qu04: Better content. As we say in ancient Rome "non ornare spirent". If there was some sort of narrative or thematic motivation to the cuts they would be better cuts. And if the music didn't suck so hard it would help too.

As far as I could see there was some sort of story involving the bored looking woman getting out of bed, going to the bathroom, lighting up, ignoring the indy rock dudes, sitting on the couch, looking in the fridge for booze, with the grand finale of wandering outside looking a bit like she was going to chuck. But then she didn't. This is NOT really enough content to make a three minute video. Waving a camera around while nothing is happening and pretending it's a film is the kind of thing preschoolers do when their parents let them play with their iphone. Only when my

kids do it there\`s not usually a bunch of smelly indy-rock dudes hanging out in the lounge room.

Qu05: I don't know what this question means. How would the editing affect the audience? Well you\`ll probably get a better reception than if you just play the unedited rushes.

Participant 3

Gender: male

Age group: 31-40

Experience: 1-3

Qu01: quite impressive, but lacking in meaning. Different versions gave more weight to different parts of the scene however this was seemingly random and some of the randomness disrupted my attention.

Qu02: At first glance the timing of cuts work ok, however there is little differentiation of the *best* part of the shot to use and the duration of the shot. Some sections were poorly paced in that the cuts went with the music for the most part but at times the music shifted gear and the editing did not. Also vice versa some fast cuts were ill-timed/ not called for in my opinion.

Qu03: Cutting between model and band in the final chorus. However there were some repeated shots and jump cuts that didnt make sense to me.

Qu04: Maybe have a similarity detection to avoid jump cuts between similar shots. Do not repeat overlay shots. If one shot has an action in it - Do not cutaway for several seconds and then cut back to it at the same time we left - ie time seems to have stood still in that one shot while other actions happened. (girls sips and reaches to put down cup / band plays 3 sec/ girl puts cup down) <<felt odd to me.

Qu05: I think (hope) certain things would make them disengage but perhaps the audience won't notice. I feel like it had an erratic feel which may work for this song but not all. Even if going for an '\erratic feel\' this could have been amplified with better shot choices and pacing. Very impressive but sometimes meaningless (for now)

Participant 4

Gender: male

Age group: 41-50

Experience: 9-15

Qu01: Fascinating to see 3 subtly different takes on the same topic.

Qu02: It was all quite choppy (in a good way) - taking simple locations and setups and adding staccato editing treatments.

Qu03: In all three there were some elements where there was subtle time shifting going on within one movement (eg drinking the bottle of wine) - it's raised to her lips, then it's back down again, then it's at her lips again for a sip

Qu04: Closeup lip syncing of the singer that was out of sync is an issue - it's quite noticeable - it didn't happen all the time mind you but when it did it was quite noticeable and led to the impression something was 'wrong'

Qu05: Unless one is trying to tell a very specific story (eg a 'standard' 'have the artist mouthing the words in a few locations / outfits with random cutaways') kind of clip this would be perfect. If it was a very specific narrative though I don't think it would work.

Participant 5

Gender: female

Age group: 21-25

Experience: 1-3

Qu01: I liked all of them. Some were a bit more 'jerky' with the jump cuts then others but that's the feel of the video so it works. The storyline wasn't effected by through out the different versions. I did like that in some versions there were more repeated shots than others.

Qu02: I love the washed out feeling of the video. It makes me feel like I am drunk. I loved the POV shots. The sharp jump cuts were also really effective in making this video really edgy. The close ups on the bands faces were also nice as a cut away between 'Eloise's' story.

Qu03: I loved the end 'chase' scene when she is running through the trees and down the road. The jump cutting with the beat of the music was really great. I also liked the grossness of the spaghetti eating. The washed out colour made it really pop in frame.

Qu04: In some shots it seems that the singer is out of sync with the music. That might just be video lag though. Continuity as well, I never saw the girl put up her hair, and then at the end it is in a bun. But that is not a HUGE deal. Qu05: I'm not really sure how it will effect it. I like they would like it. Its edgy and suits the song.

Participant 6

Gender: female

Age group: 41-50

Experience: 9-15

Qu01: the first and third had lip synching issues, which are ok in music vid but best not for close ups.

, apart from the continuity being bad in the middle. She didn't take Oyster Bay from the fridge, but was drinking it. That would have been ok if it felt like she was drinking all day, but to me it felt like she'd had a big night and was drinking in the morning.

the flashback to the cigarette butts in the bottle was unnecessary in the third one.

Qu02: jumpy, repetitive

Qu03: in some ways it gave me a giddy feeling so I felt a bit like her. I liked when he was singing/yelling at her.

Qu04: Vastly. No repeats, tell the story, don't jump backwards near the end

Qu05: it makes it unlikeable. I wouldn't want to watch it again

Participant 7

Gender: male

Age group: 31-40

Experience: 4-8

Qu01: Overall, the pacing seemed similar between them, although the first was possibly the slower version. Each had a different mix of priorities with the band vs. the girl, as well as the lead singer vs. the rest of the band.

Qu02: There was obviously a progression in the story. Some seemed to focus more on different aspects of the story - one on the food, one on the alcohol, one on the couch (weirdly...).

Qu03: In general the pacing was appropriate, although one of the cuts got too choppy at the end. I preferred the ending in the first video, which was just a long shot of the girl outside. Wasn't quite sure what to make of the shot between bottles in the other two.

Qu04: The lip syncing was very hit-and-miss. Quite distracting.

Qu05: I'm not quite sure what you're asking? Would each of the different cuts make a difference to the audience? Possibly, but it was pretty similar really, based on the footage that was planned and shot.