Marshall Plan Scholarship Research Paper

„Cutting Rhythms in Contemporary Music Videos and its Possibility of Automation“

Research conducted at San Diego State University, San Diego, CA, USA

Fall 2011

by
Isabelle Janda
University of Applied Science St. Poelten

Abstract

Rhythm is an essential and fundamental part in shaping a film or video edit. It influences the way how audiovisual media is perceived on a universally understandable but subliminal level, and it plays an important part in emotionalizing its content. Although designing the rhythm of cutting is often described as an intuitive matter we can find certain theory behind it. Beyond artistic expression it can fulfill particular communication and dramaturgical tasks and it has certainly a notable effect on the outcome if not shaped properly. Especially in the field of music videos, where we can find a minor narrative approach, cutting rhythm has significant importance in building up atmosphere and tension and should be purposely arranged and considered.

But what if these meaningful compositions are no longer made by human beings but by computer software that automatically generates the edit of the next popular music video. Especially in the last decade technical improvements and extensive research has provided the basis for intelligent automation algorithms tracing rhythmical music patterns, image semantics and other kinds of features, believed to be valuable, to imitate the human approach. Many of them can already been found in consumer editing software products enabling users to automatically arrange their home video clips to music and some of the most recent research papers in the field of computer science even show promising outcomes in the fields of narrative editing. But how effective are these algorithms and can they be applied to professional and commercial purposes?

Table of Contents

ABSTRACT **2**

TABLE OF CONTENTS **3**

1. INTRODUCTION **4**

2. CUTTING RHYTHMS **4**

2.1. RHYTHM – THE PULSE OF LIFE 5

2.2. RHYTHM IN AUDIOVISUAL MEDIA 6

2.2.1. Cutting Rhythm in a Psychological and Neurological Context 7

2.2.2. Rhythm in Narrative vs. Non-narrative Audiovisual Media 11

2.3. RHYTHM AND EDITING TECHNIQUES IN CONTEMPORARY MUSIC VIDEOS 11

2.3.2. Film vs. Music Video Editing Techniques 12

2.3.3. Cutting Rhythms and Music/Sound/Style 15

3. AUTOMATION POSSIBILITIES IN CONTEMPORARY MUSIC VIDEOS **17**

3.1. THEORETICAL DISCUSSION & RESEARCH APPROACH 17

3.2. RESEARCH RESULTS AND PRACTICAL USE OF CUTTING AUTOMATION 19

3.2.1. Creating Music Videos using Automatic Media Analysis 19

3.2.2. Creating Music Videos using Automatic Media Analysis 22

3.2.3. Automated Music Video Generation Using Multi-Level Feature-Based Segmentation 26

3.2.4. Mining Association Patterns between Music and Video Clips in Professional MTV 28

3.2.5. Web-based Approaches for Music Video Automation 33

3.2.6. Existing Music Video Automation Software 34

4. CONCLUSION **34**

BIBLIOGRAPHY **35**

TABLE OF FIGURES: **37**

1. Introduction

.

While I researched for this paper I realized that cutting rhythms and automation is not an easy scientific topic to experiment with. Not only due to their complexity but because it is so difficult to objectively evaluate results and because there are so many different approaches. Though I discovered that it is a rather complex area I found it extremely interesting and promising, also because of the fact that there is no existing general and comprehensive approach of combining the few we know about cutting rhythm with existing technical automation methods. Therefor this paper should serve as a first attempt to understand the all-embracing connection of rhythm and its effect on audiovisual media and apply observations to practical examples and conducted research. I will therefor try to explore the possibilities of non-narrative automation by using both technical and cognitive approaches. I will especially focus on automation by rhythmic music patterns and therefor analyze and compare different outcomes of existing approaches. The first part of my work will cover a theoretical discussion on cutting rhythms and their meaning for narrative and non-narrative audiovisual media. As a next step, I will try to bring them into context with musical rhythms and analyze their impacts on each other. In the second chapter the latest researches on automation in the audiovisual sector will be introduced and I will give some examples of their results. I will examine their development and try to evaluate them by applying rhythmic knowledge. Finally the last chapter will contain my conclusion and findings of my research which hopefully eventually gives me the opportunity to make some suggestions about further improvements and to state my opinion about its possible significance for the future.

2. Cutting Rhythms

To understand and evaluate the importance of rhythm in the process of video editing, I believe we first have to discover the possible general meaning and extend of rhythm for human beings in every day’s life.

2.1. Rhythm – the Pulse of Life

“What is rhythm?” – This is actually a rarely asked question because everybody already knows what rhythm is - at least everybody knows how rhythm feels, because we can experience rhythm every day. Rhythm surrounds us and not only in terms of music! Whether it is a mechanical or a natural process taking place - almost everything in our world can be described in a rhythmic pattern. Seasons, tides, days, months, years, and the movement of the stars are all examples of universal rhythms and we ourselves are functioning as a part of it. It is a rhythmic universe designed by physical laws and our survival depends on oscillation with these rhythms.[1]

By definition rhythm can be described as a “movement marked by the regulated succession of strong and weak elements, or of opposite or different conditions”[2]. And as we can physically perceive all kinds of movement with our human senses, rhythm influences us from the moment we are born.

While there are many external rhythms we sense in our lives, we also follow internal rhythms that affect us individually. One of the most dictating and vital rhythm in a human life is the human heartbeat. Though it might be in constant change, always adapting to its body’s needs, it is one of the most important factors that give us a sense of time and pace. Already as a fetus we experience our mother’s heartbeat as an all surrounding and life-dictating pulse. This pulse is not only the very first thing we perceive with our human senses there are even recent studies that the fetus aligns its own heartbeat to the one of its mother.[3] It is also tightly linked to our bio rhythmical balance. Waking/sleeping, eating/digesting, working/resting, and inhaling/exhaling are just examples of action opposed to each other that form our daily routines to survive in this world.[4] But we also create rhythms for other purposes. They give our lives structure and they even help us to communicate. We use rhythms in language to give words different meanings or emphasize the emotion behind it and of course we use music as a rhythmic tool to express ourselves artistically. Although the individual interpretation of rhythm can differ from each other it can be said that rhythm itself can be universally understood and is perceived on a physical level by every living being. As we can clearly see, rhythm is what

[[1]] Pearlman, 2009, p. 7

[[2]]The Compact Edition of the Oxford English Dictionary. II., p. 2537.

[[3]] BBC News, 07/20/2010, „Mother’s heartbeat ‘synchronises with foetus’”

[[4]] Pearlman, 2009, p. 7

moves us through life and it can be experienced in all kind of ways, playing a crucial role in our world.

2.2. Rhythm in Audiovisual Media

Probably everybody knows the feeling when watching *“a well-crafted scene in a movie, documentary – even an online video. The world outside melts away and our internal energy and emotion are in sync with the energy of what we're viewing*.”[5] Some videos “*grab us, pull us in and capture our imagination for the few minutes we watch them*”[5] But why do some pieces of audiovisual media have such an effect on us while others just pass us untouched. The answer for this is of course not a simple one and can be traced back to several factors. But although the experience and perception is highly individual and different preference might influence our judgment, a general consensus concerning the media’s quality can be seen among the audience. Our long-time experience shows that the audience attitude towards the product can tightly be linked to a feeling of a “right” rhythm. This rhythmic knowledge is learned over time[6] and is not only limited to the frequency of cuts, the physical rhythm but in strong connection to the content itself concerning emotional and event rhythm.[7] Although hard to articulate we seem to be drawn to an overall pulse that, especially when not shaped right, catches our attention and has an deep impact of whether or not we can emotionally connect with the seen.

One of the very few and pioneering pieces of literature that addresses this topic for itself is “Cutting Rhythms – Shaping the Film Edit” by Karen Pearlman. In her doctoral thesis Pearlman tries to articulate underlying principles of what rhythm in film is and connects theory and practice in new and inspiring way. Derived from her studies and her own experience as an editor, she proposes that “*rhythm in film editing is time, movement, and energy shaped by timing, pacing, and trajectory phrasing for the purpose of creating cycles of tension and release*.”[8] – A profound knowledge on which I will base my research on and explain in some detail later on.

[5] Hayduk, Kelly, 10/10/2011, „How Rhythm Keep Us Watching“

[6] Pearlman, 2009, Introduction p. xxi - xxiii

[7] Pearlman, 2009, p. 84-87

[8] Pearlman, 2009, Introduction p. xxiii

As she, as well as many other editors, point out, rhythm in audiovisual media seems to be an “individual and intuitive matter”[9] and, because of that, has not been researched thoroughly on a theoretical and cognitive basis. Nevertheless we can find strong indicators that along with intuition shaping rhythm can be learned and used deliberately to create different outcomes. In the following chapters I will, therefor, try to extract important features of this research that may be of importance to the question, whether or not a complex process like this can even considered to be automated. And although it should be said that of course “cinematic rhythm as a whole derives not only from editing but from other film techniques as well”[10] it seems that editing is one of the most significant tools for doing so.

2.2.1. Cutting Rhythm in a Psychological and Neurological Context

So one of the most interesting questions in this context is how are human beings capable of perceiving rhythms neurologically and even creating themselves as well as how cuts work to create a cinematic rhythm. Walter Murch, one of the most experienced Hollywood film cutters, introduces his own, but very interesting, theory about the effect of cutting in his book “In the Blink of an Eye”. There he suggests that, although cuts seem to be an artificial device created by human hand and “is not anything we experience in ordinary life”[11], we might carry a similar habit that helps us to understand why we can deal with the instantaneous discontinuity of cuts the way we do. In his book he clearly explains why the physiological mechanism of eye blinking almost seems to be a paradigm for the way we cut and perceive cuts. His theory becomes even clearer when explained through a practical example by John Hustler:

“*Look at the lamp across the room. Now look back at me. Look back at the lamp. Now look back at me again. Do you see what you did? You blinked. Those are cuts. After the first look, you know that there’s no reason to pan continuously from me to the lamp because you know what’s in between. Your mind cut the scene. First you behold the lamp. Cut. The you behold me”*[12]

[9] Dancyger, 1997, p.307-315

[10] Bordwell, Thompson, 1997, p. 278-280

[11] Murch, 2001, p.57

[12] Christian Science Monitor, August 11, 1973. John Huston interviewed by Louise Sweeney.

And furthermore Murch provides us with an interesting statement about rhythm in this context:

"*So it seems to me that our rate of blinking is somehow geared more to our emotional state and to the nature and frequency of our thoughts that to the atmospheric environment we happen to find ourselves in… If it is true that our rates and rhythms of blinking refer directly to the rhythm and sequence of our inner emotions and thoughts, those rates and rhythms are insights into our inner selves*"

Finally he concludes that "*these juxtapositions are not accidental mental artifacts but part of the method we use to make sense of the world: We must render visual reality discontinuous, otherwise perceived reality would resemble an almost incomprehensible string of letters without word separation or punctuation. When we sit in the dark theater, then we find edited film a (surprisingly) familiar experience*."[13]

Murch has given us a first basic approach of how we perceive cutting rhythms, but how does an editor actually gets to the decision of making a cut at the right time? We already know that it has something to do with intuition. But how is intuition actually developed or acquired, and is it something people are born with? Again Karen Pearlman has found an adequate answer for this topic and adapts Guy Claxtons 6 different types of intuitive thinking for film editing. These types of thinking are believed to be at work in different combinations when we speak of intuitive editing. Those are the following:[14]

1. Expertise – the unreflective execution of intricate skilled performance
   "*It's a matter of knowing your gear of choice so expertly that its operation doesn't require conscious thought*"

2. Implicit learning – the acquisition of such expertise by non-conscious or non-conceptual means
   "*There are conventions of filmmaking that show up in most TV programs, ads, … An editor may not know the names of these conventions or techniques but has seen them enough to know what they are without ever having consciously learned them*."

[13] Murch, 2001, p. 62-64

[14] Pearlman, 2009, p. 3-5

3. Judgment – making accurate decisions and categorizations without, at the time, being able to justify them
   "*Judgment can be seen at work whenever an editor makes and adjustment to a cut and it works better. Once the 'working better' is visible, an editor is rarely called upon to explain why or how*."

4. Sensitivity - a heightened attentiveness, both conscious and non-conscious, to details of situation
   "*An editor has sensitivity or heightened attentiveness to movement and emotion in the material*"

5. Creativity – the use of incubation and reverie to enhance problem solving
   "*Editing creativity is the lateral association of images or sounds to solve the problem at hand, which is the shaping of the film and its rhythms."*

6. Rumination – the process of "chewing the cud" of experience in order to extract its meanings and it implications
   "It is the kind of thinking when you're thinking about something else, and you have immersed yourself so deeply in your material that it inhabits a part of your brain even when you're not actually looking at it or working on it.

Summarizing and concluding the above: "*Intuition isn't something you just have. It is something that can be developed, enhanced and even acquired through practical and theoretical experience and education*."[15]

So how do we actually acquire rhythmical knowledge neurologically? There are at least two physiological activities which are believed to be crucial for this process. "The first one is 'kinesthetic empathy'. Kinesthetic empathy is feeling with movement, a sensitivity we have developed by perceiving and being movement"[16] Neuropsychologist Arnold Modell

[15] Pearlman, 2009, p.6

[16] Pearlman, 2009, p. 10-11

describes the activation of kinesthetic empathy by saying "The perception of feelings relies on the corporeal imagination which in turn is determined by the history of the self"[17]

In other words our body tries to imagine, from its own experience, how another body feels. And it does so, surprisingly, even if it has not actually experienced this kind of movement. "…if we have not fallen in a fast, straight, hurtling trajectory, our bodies know to duck if something comes hurtling at them, just as they know to brace for impact if they themselves are falling….So, movement speeds, directions, and energies have meaning when we see them, even if we have not experienced them"[18]

The second special feature of our brains is what neurologists call "mirror neurons". Mirror neurons were first detected in monkeys and are neurons that "*discharge both when a monkey performs certain movements"* but as well "*when the animal merely observes another monkey performing the movement*."[19] This is believed to be the case in human brains as well and is such an remarkable discovery because it may be the key for understanding why humans can comprehend the actions of other people and how the learn new skills by imitation.[20]

To return to our topic of cutting rhythms, this might be one of the reasons why audience senses the same kind of rhythm in audiovisual media and empathize with the illusion of cinema. "*Mirror neurons allow us to participate in another person's intention movements. Our neurons do the movement with them, whether they are live or on the movie screen*."[[21]]

And in a similar way this could explain the working process of an editor.

"*So, what an editor may be doing in making rhythm in moving pictures is engaging her corporeal memory and/or mirroring, neurologically, part of what she sees and hears*"[22]

But by using these native rhythm building blocks something new evolves.

"*Putting two shots together, each of which inherently has rhythm, makes a third rhythm, which is not the same, or even just the sum of the first two. So the edit begins to have a rhythm of its own*"[22]

[[17]] Modell, 2003, p.145

[[18]] Pearlman, 2009, p.11

[[19]] Restak,2003 , pp. 35-37

[[20]] Ramachandran, V.S., "Mirror Neurons and imitation learning as the driving force behind "the great leap forward" in human evolution". Edge Foundation. Retrieved 2006-11-16.

[[21]] Pearlman, 2009, p.12

[[22]] Pearlman, 2009, p.13

This also explains why editors produce different, creative and unique outcomes although editing the same material.

"*As editors begin to do more than neurologically imitate existing rhythms, they draw on rhythms inside themselves, as well as those things captured in the rushes, to create the film's rhythm*.[22]

2.2.2. Rhythm in Narrative vs. Non-narrative Audiovisual Media

Up to this point, I referred to all kind of cutting rhythms when speaking about audiovisual media but in fact there is a huge difference between a narrative and a non-narrative approach. While the narrative approach concerning rhythm seems to have a more complex structure by including content-based semantics and is very much influenced by story-telling techniques and its tension & release cycles (see Pearlman, 2006, Chapter 4), the non-narrative approach can be deemed less strict and more intuitive in its assembly. Especially music videos, which certainly have a minor narrative approach, are mainly referred to as audio-oriented and therefor have to be treated differently. This being the case, my paper will only focuses on a non-narrative approach and will only conduct research leading toward a non-narrative outcome.

2.3. Rhythm and Editing Techniques in Contemporary Music Videos

Since the 1980's, when music videos, traditionally called MTV - Music TV, became very popular for the first time, we know them as rich visual representations of popular songs in TV or other mediums. Although music videos are primarily made and used as a marketing device, time showed that this visual representation can be rather artistically and add value to the otherwise independent existence of the song itself. Furthermore, especially for the young generation, it has become a "representative art form of our new media age"[23] determining values and setting trends.

But what exactly defines a music video in terms of editing and looks. This is actually hard to say because there are no certain rules of how a music video has to look like or has to be edited.

[[23]] http://www.aboutvideoediting.com/tutorials/music-video.shtml

"*Music videos use a wide range of styles of film making techniques, including animation, live action filming, documentaries, and non-narrative approaches such as abstract film. Some music videos blend different styles, such as animation and live action. Many music videos do not interpret images from the song's lyrics, making it less literal than expected. Other music videos may be without a set concept, being merely a filmed version of the song's live performance*"[24]

And not only this, furthermore it is in constant change and one of the fast developing genres of our times. As Daniel Moller points out in his paper "Redefining Music Video", that exactly centers on this particular question, "the way we define Music Video has not kept pace with the genre's evolution"[25]. Not only are most of definitions not able to include the vast "abstraction, motion graphics and experimental nature of many music videos"[26], even the aim of serving as a marketing device for a song has shifted to something else.

Concluding this, it can be stated that music videos have very few appearance conventions developed, their use can actually range from advertisement to artistic purposes and it is "*one of the most active, fast changing types of videos with most varieties*." [27]

2.3.2. Film vs. Music Video Editing Techniques

We can see, that it is very difficult to define contemporary music videos but while we are not clear about a music video stereo type we can, however, find similar characteristics that help us determine an appropriate editing rhythm and editing techniques. First of all we have to state that music videos significantly differ from any other kind of video type because "they are more audio-oriented and the visual aspect of the videos can be considered secondary[28]". This of course only applies in terms of content meaning and not in terms of style, which can be seen as one of the most significant aspect of a music video.

"The most obvious characteristic of music video is that it's highly stylized.... *you may not be able to recognize the screen time and place, you may not be able to tell what's*

[24] Cutietta, 1985, "Using Rock Music Videos to Your Advantage". Music Educators Journal 71 (6): 47–49

[25] Moller, 2011, p.1

[26] Moller, 2011, pp.1-8

[27] http://www.aboutvideoediting.com/tutorials/music-video.shtml

[28] Sosongko, 2011, p. 19

happening on the screen. But there's one thing you will feel unavoidably: its style, its flavor"[29]

The second major characteristic of any kind of music video is its aim to evoke a certain feeling.

"…what's important for music video is not logic, or story, it's the feeling and mood."[29]

So instead of narrative goals the MTV has replaced it with a multilayered approach where the place, feeling or mood can be seen as the primary layer.[30] "*All the short plots may happen at different places, consist different characters, in different seasons, but they all contribute to one single feeling*."[29]

These important characteristics now help us to understand why music videos are edited so much differently than other kinds of audiovisual media and if they could even qualify for automation.

But what do we mean in particular if we speak about "MTV-style editing"? As we compare this style with film editing, we certainly think about a more rhythmic but disjunctive matter. We might refer to music editing as quick cutting or editing on the beat and we can see that cuts appear more frequently than in a regular movie. Concluding that our shaping device is music, narrative is less important and the feeling state is our primary layer we expect a higher use of jump cuts rather than match cuts.[31] But especially because of its dependence on music, music video editing, much more than classical Hollywood film editing, bears a responsibility for many elements:[32] While film editing must maintain spatial and temporal continuity, disguise cuts, employs cross-cutting to build suspense or create thematic associations, and so on, music video editing does all this, but just as often it does the opposite.[33] Music editing can direct the flow of the narrative; it can underscore non-narrative visual structures and form such structures on its own and, like in film, it can color our outstanding of characters but because of its responsiveness to the music it insures that no single element – the narrative, the setting, the performance, the star, the lyrics, the song – gains the upper hand.[34]

[29] http://www.aboutvideoediting.com/tutorials/music-video.shtml

[30] Danzyger, 2011, p. 165

[31] Danzyger, 2011, p. 166

[32] Beebe, Middleton, 2007, p. 125

[33] Beebe, Middleton, 2007, p. 146

[34] Beebe, Middleton, 2007, p. 125

Going into detail we can see that for some parts music video editing follows traditional narrative film practices, yet it also extends and breaks these rules. This is most notably in the use of continuity editing, where we can see that many music videos constantly switch from short match cut phases to immediate disjunctive edits. These techniques often cause a certain feeling and effect on the audience but furthermore determine a certain rhythm.

“Music videos avoid matches on actions, often extending or abbreviating a shot to give the sense of a cut in the ‘wrong place’. This effect blunts narrative progress and creates a rhythmic emphasis on the moment when the edit occurs”[35] It draws our attention to the cut itself and makes us think of how this two disjunctive shots relate to each other. “It also implies a centrality for pace…Consequently, pace becomes that source of energy, and new juxtapositions that suggest anarchy and inventiveness.”[36]

As Carrol Vernallis points out in his observations on music videos:

“The power of music video editing comes from the fact that there is no predicting what a video’s edit will do. An individual edit in music video can also carry great weight, perhaps more than cinema. In music video, shifting a shot by a single frame can dramatically alter the feel of the tape: it can determine the way that subsequent shots relate to earlier ones, and it can also shape what the viewer notices in the music”[33]

Another major difference between music video editing and film editing is that more 95% of all shots in typical music videos are moving shots. If not the camera itself is moving, the objects within are in constant change. As we already know we perceive rhythm as movement this inner frame rhythm has particular meaning for the look and feel of the overall outcome.[37]

So although it seems chaotic, music videos bear a structure derived from its rhythm. This rhythm is partly formed by “the way that editing is sometimes very noticeable yet sometimes invisible as it draws from the techniques most common in film”[34] but is also very influenced by the music accompanied, as I will treat in the next chapter.

As we can see this topic is so complex that it would go beyond the scope of this paper to go into all the little details and cutting techniques any further, but it is now clear that when we think about editing automation we have to consider all characteristics and techniques to make a music video as rich as possible.

[[35]] Beebe, Middleton, 2007, p. 127

[[36]] Danzyger, 2011, p. 166

[[37]] http://www.aboutvideoediting.com/tutorials/music-video.shtml

2.3.3. Cutting Rhythms and Music/Sound/Style

In this last paragraph, before eventually treating the topic of automation, I shortly want to explore the very important relationships of cutting rhythms and music respectively sound, as well as style in terms of a particular theme or feeling. These relations, on the one hand, seem very obvious to effect cutting rhythms, but on the other hand they are very hard to detect and determine. Though I could only find little theoretical discussion about their interaction I believe them to be crucial for the question of possible editing automation and I will therefor refer to this thesis in my further studies frequently.

Naturally, music and sound are the carrying element of any kind of music video. While in film editing sound and music are supposed to underline, counterpoint, enhance, contradict, shade, etc., the pictures[38], in MTV it is exact the other way around. As pictures must support and emphasize the music it of course implies that music rhythm is a stronger leading element in music video editing.

Editing, in general, is in fact often compared with the craft of music making and is particularly made analogous to the process of composing, orchestrating and conducting. Though these terms refer to the cutting process in some way, they are not particularly precise comparisons. Editors for example, do not exactly make up their cutting rhythms like composers their musical rhythms but in fact assemble them by choosing different selections, orders and durations of shots.[39] This analogy has though particular meaning in the area of music video, where pictures and its space resp. objects become instruments of rhythms itself. Because rhythm is perceived by movement we can detect that “the cinematic features and mis-en-scène of music video – extreme high, low and canted angles, long tracking shots, unusual camera pans and tilts, and the lively features within the frame – can mimic sonic process.”[40] Following this we can even say that “the types of shots used in videos do not just reflect sonic processes but also suggest a listening subject as much as a viewing one.”[40] Vernallis therefore concludes that watching music videos create an experience more like listening than viewing.

[[38]] Pearlman, 2009, Introduction p. xxvi

[[39]] Pearlman, 2009, pp. 24-25

[[40]] Beebe, Middleton, 2007, pp. 131-132

Another factor in this relationship is that "music – particularly without lyrics - synthesizes human emotions."[41] As we already know, one of the most important characteristics of music videos is its aim to provoke a feeling and so the visual elements try to heighten this emotional experience with adjusting to the music. This feeling state can be sharp and deep or can be developmental and dreamlike, either way it gives us a certain disjunctive, disconnected sense of rhythm. If we add lyrics we are given a direction for this emotional state and a third rhythmic component comes into play. Especially in Hip Hop and Rap music language rhythms has to be taken into account when determining a cutting rhythm. The combination of these three rhythmic elements (music, editing & language) can be seen as the basis of the overall rhythm we perceive while watching a music video. Enhanced by the movement within the frame and our own inside rhythm we start to realize that this topic is a rather complex one.

We can see the relationship between music rhythms and cutting rhythms is very strong and crucial but how they actually interact with each other remains to be researched. For this reason the last chapter of paper will especially focus on experimenting with these interactions and will hopefully help us to clarify its significance for the future of editing automation.

Finally I want to draw the attention to what we call "style" and its impact on the rhythmic appearance of MTVs. As with all areas of editing the word "style" refers to an aggregate of choices (what shots, in which pace, what kind of digital effects and so on).[42] While we typically refer to "music video editing style" in general as a characteristic of music videos we can also understand "style" in an extend way to describe the overall idea, the "look and feel" or the emotional state it conveys. Although a very amorphous term it is what every music video aims to achieve and it is what should be left with the audience after watching it. Style is what separates music videos from each other and it is the maintaining of style over the duration of an edit that makes it so interesting for the question of rhythm and eventually for the question of automation.

It is the choice of style that eventually makes every single music video unique in terms of editing.

[41] Danzyger, 2011, p. 168

[42] Pearlman, 2009, p. 153

3. Automation Possibilities in Contemporary Music Videos

Dealing with the matter of automation one must always also deal with question of aesthetic and artistic expression and even with the question of sensuousness of this research process. Though these questions are very important and should always be asked along a research I will not go further to challenge this until I've examined some of the most recent results in research. I assume that some of these questions will resolve when seeing the final outcome and it will provide a more stable basis to argue with.

There are some questions though that should be asked before stating any opinion and that are fundamental for any scientific approach. Therefor in the next chapter will contain a small theoretical discussion, likely rising more questions than answering, but providing a base of what to expect from this research. I will try to find criteria how to measure the quality of results and I will include my observations from the chapters above. In the last paragraph of this chapter I will eventually introduce and comment some of the recent research papers, automation software algorithms and their results concerning editing automation.

3.1. Theoretical Discussion & Research Approach

Editing automation has become more and more interesting in the last few years. Not only because technical improvements have provided better results and new possibilities but also because the demand of home video cutting and audio-visual consumer programs has risen immensely. In a digital age like this consumer become producers and the diversity of youTube and other online media platforms show how effective and popular home editing has become for entertainment purposes. While quality becomes secondary, editing evolves more and more from a profession to a standard creation tool for everybody. That is why software developers gain more and more interesting in algorithms that help simplify this process for their users. This enforces research and over the last decade a lot papers have been published confirming this development:

"*We believe that the combination of high-quality digital video, cheap capture cards, low cost disk space, and interest in creating video content for the Web will increase the demand for editors that handle non-professional video material*."[43] – Opening Words of the Paper "*A Semi-automatic Approach to Home Video Editing" published in 2000*

[43] Girgensohn, Boreczky, Chiu, Doherty, Foote, Golovchinsky, Uchihashi, Wilcox, 2000, p.81

"Digital video has become affordable and attractive to home users, but skill and manual labour are still required to transform amateur footage into aesthetically pleasing movies."[44]
– Introduction of the Paper "An Evolutionary Approach to Automatic Video Editing" published 2009

But how much can we even expect from this development? As we leave the editing process to a machine we certainly have to ask ourselves how much "quality" can it produce and can it achieve the same effects on human beings as something that was created by man?

The question of quality is of course one of the most important ones when evaluating the outcome of computer generated videos. But how can we even measure and compare quality in this difficult and complex area. As we already know from above our sense of quality has something to do with the rhythm transferred by the video and our own subjective perception. Though we cannot measure the impact of individual perception by cognitive approaches we have to leave this aspect to psychological researches and survey instruments. What we can do here is to compare the features we believe to be valuable for our sense of rhythm, such as physical, emotional and event rhythm. We can compare the interaction of music and cutting rhythms and we can extract tension and release cycles within the overall rhythm. Nevertheless in music videos we can only truly judge about quality by checking if it has achieved its primary goal - to provoke a certain feeling.

Knowing this, in the next chapter, the first practical part of my research, I will treat some of the most interesting papers on automatic generation of music videos and I will try to comment their results by comparing them with our cognitive approaches and extracted rhythm features from above. I expect this to give us a profound overview of how far research has gotten to this particular point. The second and final part of my research will then experiment with actual music video examples. This should show how the rhythms differ in automated and handmade music videos by analyzing and comparing rhythmic features. Through this research approach I hope to find an answer to my research question of how cutting rhythms work in music videos and if automation has future for professional purposes.

[44] Wang, Mansfield, Hu, Collomosse, 2009, p. 1

3.2. Research Results and Practical Use of Cutting Automation

One of first and pioneering papers I found on the topic of automated music video generation was “Creating Music Videos using Automatic Media Analysis” publish in 2002 by Jonathan Foote, Matthew Cooper, and Andreas Girgensohn.

3.2.1. Creating Music Videos using Automatic Media Analysis

Conducted in FX Palo Alto Laboratories, a leading multimedia research laboratory located in Silicon Valley, this research group claims to have developed methods for automatic and semi-automatic creation of music videos superior to any existing work before. Their outcome: “*a fully- or semi-automatic video summarizer that con condense a lengthy home movie into a compelling 3 minute music video*.”[45] - Not exactly suitable for professional purposes but a first innovative approach towards editing automation by extraction of audio and video features.

Their working practice: significant audio changes are automatically detected from a given arbitrary audio soundtrack and recorded during the entire duration of the music track. Similar to this the source video is automatically segmented and analyzed for suitability based on camera motion and exposure (“*video with excessive camera motion or poor contrast is penalized with a high unsuitability score*”[45]). High quality video clips are then automatically selected and aligned in time with significant audio changes by choosing the most suitable region of the desired length. Besides this fully automatic process, clips can also be manually selected and ordered using a graphical interface (“Hitchcock”).

[45] Foote, Cooper, Girgensohn, 2002, p. 1

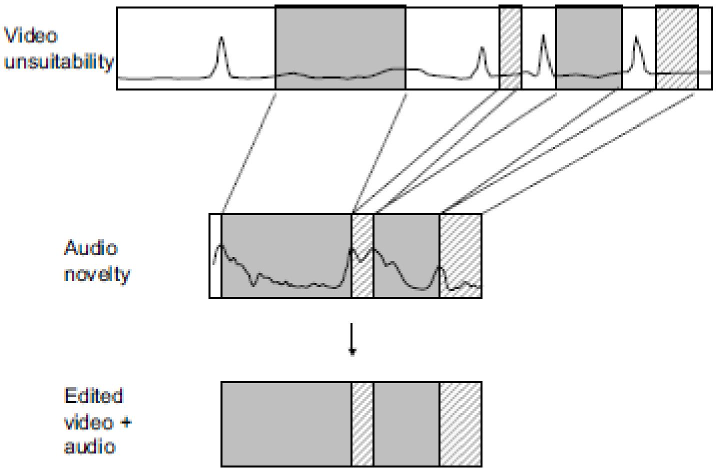

Figure 1: Automatically editing video to synchronize with a shorter audio track

The Working Process in Detail:

First video and audio are selected for input to the system. Simultaneously video and audio are then analyzed and evaluated.

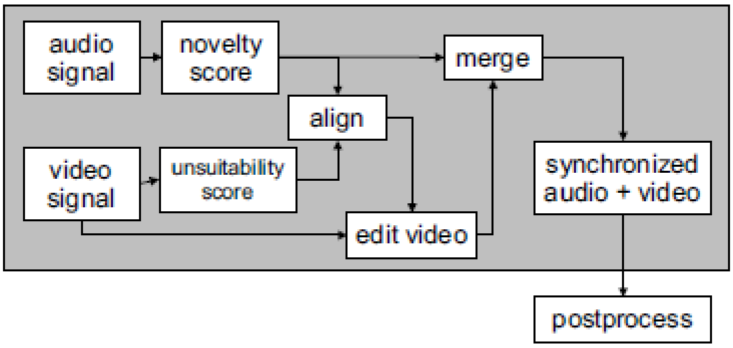

Figure 2: Automatic music video system block diagram

The source video is divided into clips, separated by areas of fast camera motion and the amount of brightness. To quantify the suitability and determine appropriate clip boundaries, a numeric unsuitability score is computed. First the amount of camera motion is detected and quantified and then the overall frame exposure is added to the score. The score peaks mark potential clip separations which then can be adjusted according to window size and peak thresholds resulting in more or fewer clips.

The audio soundtrack is initially parameterized and the examined by the use of a self-similarity analysis "*At each instant, the self-similarity for past and future regions is computed as well as the cross-similarity between the past and the future. A significant* *novel point will lie between regions of high self-similarity. These regions before and after the novel point will also exhibit low cross-similarity*"[46] If the analysis fails to detect significant changes the perceived quality of the music video will suffer eventually. In the optimal case novelty is measured as a function of time by detecting the checkerboard-like features along the main diagonal of the similarity matrix. Peaks in the novelty measure again correspond to audio segment boundaries and can be used as edit points.

Given the peaks of the audio novelty score and the video clip boundaries audio and video has to be synchronized. Therefor major peaks in audio are aligned with video clip boundaries. The necessary clip length is found from the distance between audio novelty peaks and truncated to the length of each audio segment according to the suitability score.

Results:

As claimed by the paper several music videos have been automatically created by this algorithm using a variety of video sources and audio sources such as "The Beatles", "Vivaldi", Jazz and Dance music. Though not measurable by objective criteria subjective judgments showed that the method produces reasonably convincing music videos. However the results seem to be improved by the use of a semi-automatical approach where user could choose from which clips the computer should edit from.

Conclusion & Rhythmic Evaluation

"Creating Music Videos using Automatic Media Analysis" is one of the first automation projects that include a basic rhythmic audio approach. By analyzing significant changes in the audio soundtrack video is synchronized in a rhythmic way. What is most remarkable here is that the method of self-similarity analysis shows to be robust across a wide variety of jazz, orchestral, and popular musical genres, while other approaches usually fail due to limiting assumption about rhythm. However there are of course many other automation areas that lack these rhythmic approaches. Besides that there are no considerations on in-frame movement and event rhythms or even standard editing rules this approach also misses any constitutive structure. As the video is only aligned by assumption of its quality it cannot serve as a method for any professional use. Moreover the static structure cannot serve to produce any tension & release cycles as demanded to create an emotional thrill.

[46] Foote, Cooper, Girgensohn, 2002, p. 2

Further enhancements as stated by the paper could be made by adding rhythmic synchronization - the extraction of repetition patterns and tempo to determine a minimum clip duration.

3.2.2. Creating Music Videos using Automatic Media Analysis

The next very interesting paper I'd like to introduce here is "Automatic Music Video Generation Based on Temporal Pattern Analysis" published in 2004 by Xian-Sheng Hua, Lie Hu and Hong-Jiang Zhang. This research was conducted by Microsoft Asia and shows many improvements to the paper just presented.

Though the scope is still consumer-oriented and restricted to the use of home videos we can see a first approach towards content analysis, rhythmic methods have been extended and for the first time we can see stylizing elements like transitions instead of only plain cuts.

According to the paper, they developed an automatic music video (AMV) generation system based on the set of video and music analysis algorithms, which is able to extract temporal structures of the video and music, as well as repetitive patterns in the music. Following these observations a set of highlight segments from the raw home video footage are selected, in order to appropriately match the visual content with the aural structure and repetitive patterns.

The Working Process in Detail:

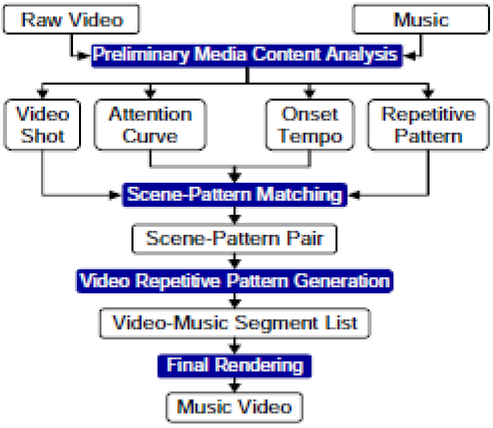

Figure 3: Flow chart of Automatic Music Video Generation

As usual music and video is simultaneously analyzed to extract valuable features.

Video Analysis:

First the raw home video is segmented into shots according to color similarities or timestamp (if it is provided or able to recognize). Then an “attention” or “importance” value is calculated for each shot by averaging the “attention index” of each video frame. This attention index is determined by the amount of object motion, camera motion, color and speech in the video.

Audio Analysis:

For estimating the musical rhythm for the whole track or single sub clips the team of Microsoft Asia applied their own, so called, “onset detector”. This algorithm, developed in one of their previous researches, extracts an onset series (“*Onset is the moment when a key is pressed down, which is used to roughly estimate the music rhythm in our system*”[47]), as well as corresponding onset strengths in the incidental music by checking “energy peaks” in frequency domains. This way tempo can be measured as a numeric value and determines cutting frequency.

In the next step the repetitive music structure of the audio track is tried to be determined. This is done by extracting a features set, including temporal features, spectral features and CQT features.

From this a structural analysis can be executed:

“Temporal features are used to estimate tempo period and the length of a musical phrase, which is used as the minimum length of a significant repetition in repeating patterns discovery and boundary termination. Spectral features are used for vocal and instrumental sounds discrimination in order to identify the ‘prelude’, ‘interlude’ and ‘coda’ of a popular song in final music structure analysis. CQT features are used to represent

[47] Hua, Lu, Zhang, 2004, p. 2

the note and melody information, based on which a self-similarity matrix of the music is obtained. The significant repeating patterns are then detected from the similarity matrix with an adaptive threshold setting method. Finally, the boundaries of repeating patterns are roughly aligned to facilitate music structure inference; and the obtained structure Is utilized correspondingly to refine the boundary of each musical section, with an optimization-based approach."[47]

After video and music is analyzed they have to be aligned. To achieve this, the paper describes three sub-steps: quality filtering, scene segmentation and scene-music pattern matching. So first low-quality shots are filtered out based on camera motion and color entropy (like we already know from the previous paper). Then shots are grouped according to content similarity and timestamp (if available). The last part is choosing which of these scenes match best to the extracted musical patterns. "*To make the final music video compelling and more like a professional edited music video, we try to match the tempos of the music repetitive patterns with the motions intensities in the corresponding video scenes, as well as to preserve the 'important' segments in the raw video*"[48]

After each music pattern is assigned with a video scene, each scene is examined for repetitive video pattern corresponding to the music patterns. If found those are extracted and matched, thus the repetitive patterns in audio track and video track are accordant. For the final assembly the research group defined 3 appearance rules to follow:

1) Transition Rule: At transitions connecting different music patterns "slow" cross-fades are applied. For other transitions, if the two consecutive sub-shots are similar in color, use cross-fades (transition length determined by strength of the onset) otherwise, a cut is applied.
2) Effect Rule: For the sub-shots associated with "prelude", "interlude" and "coda" certain transformation effects are applied, such as "Sepia Tone", "Grayscale", "Slow Motion" or "Old Movie" to enhance the feeling of pattern repetition.
3) Start/End Sub-Shot Rule: The last sub-shot of the music video is taken from the same shot where the first sub-shot is taken from. And the first sub-shot will be applied with a "Fade In" effect, and the last sub-shot will be applied with a "Fade-out" effect. In addition, captions (music name, singer's name, etc.) will be superimposed on the bottem-left corner of the first sub-shot and the last sub-shot.

[48]Hua, Lu, Zhang, 2004, p. 3

Results:

Again this group had difficulties evaluating their results, so they decided to compare their work with two other related research projects (AVE- MV style & video summarization), both projects without video-music patterns matching, and provided 15 videos in total to 10 users for subjective evaluation. Results showed that AMV was evaluated better than the other two projects. Though comparison to professional editing results have not be executed yet the outcome was promising and compelling.

Conclusion & Rhythmic Evaluation

Obviously, this paper shows immense improvements to “Creating Music Videos using Automatic Media Analysis” from 2002. While the general procedure of evaluation audio and video by a numeric value is still the same, the features extracted are more significant and also include object movement and source audio (speech) for the first time. Another huge aspect is the introduction of repetitive audio analysis. For the very first time a rhythmic approach is applied according to the musical structure and video footage is not only aligned by content similarities but also by a musical event rhythm. There is even a first attempt to apply an emotional rhythm by trying to match the tempos of the musical patterns with the camera motion of certain shots (more camera motion is usually connected to a more emotional perception). Of course it also has to be stated that this project provides just on possible solutions of many. If we look at the style guidelines we can see, that those are mostly randomly chosen and do not really follow scientific findings of the human perception. Furthermore no assumptions about shot editing rules have been made and there is no semantic content approach here. Still it is impressive work and a promising direction towards professional editing automation.

In their conclusion the group of Microsoft Asia states some possible improvements for their work I also want to mention here at last: One of their future ideas is to still preserve the storyline of raw footage while adding several repetitive video patters, corresponding to audio patterns, among them. Furthermore they propose to add interactive system, as a semi-automatic approach, which gives users editing suggestions. And another interesting idea here is to add face detection to make the output more compelling and professional.

3.2.3. Automated Music Video Generation Using Multi-Level Feature-Based Segmentation

The third and also one of the newest scientific automation project for music video is called "Automated Music Video Generation Using Multi-Level Feature-Based Segmentation" conducted by Jong-Chul Yoon, In-Kwon Lee, Siwoo Byun and published in 2008

In the abstract this research project states to promote coherent matching of audio and video by "*analyzing the flow of both music and video, and then segment them into sequence with similar flow*".[49] Segments are as usual matched according to features extracted from audio and video but in this case a multi-level segmentation approach is used to increase the structure of matching. In comparison to the previous paper this automation system does to not assemble an arbitrary sequence of video clips but instead claims to generate a music video that conserves the video-maker's intention and enhances the level of synchronization between audio and video.

The Working Process in Detail:

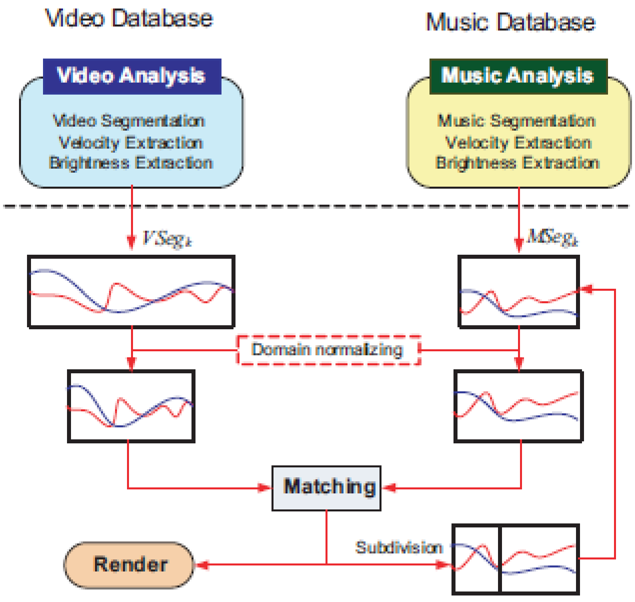

Figure 4: Overview Working Process

[49] Yoon, Lee, Byun, 2009, p.1

The system can be divided into a media analysis module and a matching module. In the media analysis tools both music and video is segmented first and analyzed for significant features.

For the music segmentation Yoon et al. suggest a novelty scoring (same as already introduced in “Creating Music Videos using Automatic Media Analysis”) which detects temporal variation of the wave signal in the frequency domain. What differs to the previous paper is how this research segments the given video files. A so called “contour shape matching” method is used to find extreme changes of shape features between frames. To support mulit-level matching each video segment is then divided into segmentation levels from coarse to fine. Then velocity and brightness features are extracted from each segment. “From the many possible features of a video, we choose velocity and brightness as the basis for synchronization. Velocity means the displacement over time derived from the camera or object movement, and brightness means the average luminance value in each frame.”50 Using these features the matching module comes into play. Because each segment of audio and video has a different length the time domain is first normalized. Then three matching terms are defined as below:

- Extreme boundary matching: In the case of video, the changes in Hu-moments determine the extremes of flow. Novelty scoring locates points of extreme variation in the music, which can be matched with the changes in the Hu-moments.
- Velocity matching: The velocity features of the video and the music can be matched to synchronize the video with the beat of the music.
- Brightness matching: The brightness features of the video and the music can be matched to synchronize the timbre of the music the brightness of the video.

Additionally to these 3 terms, two terms based on the velocity histogram and segment length are added and combined into a cost function to guide the matching. Thereby the users can change the weights to control the individual importance of the matching terms. Audio and video pairs with the lowest cost values are then matched and a time warp effect is applied to match their different lengths. To avoid repetition of video segments these selected pairs are not included in the next recalculate process. If the matching module cannot find a satisfactory match, a multi-level segmentation is applied, which further subdivides the music segment, and repeats the calculation to improve the match. To avoid uncontrolled subdivision, three levels is the maximum for this procedure.

[50]Yoon, Lee, Byun, 2009, p.3

Results:

Yoon et al. tested their system with different self-made videos and audio tracks for its performance but unfortunately the paper does not state any result testing. Though it mentions that although each segment is matched in terms of features, it is possible that points of discord still exist within a segment.

Conclusion & Rhythmic Evaluation

This paper is in particular interesting because it is the first approach that considers in-frame movement and its rhythm. While other studies "*treat video segments as primitives to be matched and do not consider the flow of the video segments*"[51], this Korean research segments video by changing shapes. Furthermore the synchronizing process is redefined through a multi-level process that keeps looking for best matches in velocity and brightness. Another new idea is that video clips get "time-warped" to fit the musical patters. If this provides a more stable overall rhythm or confuses the audience is not be argued until tested by further studies. What this approach lacks though is the innovative method of adding structure by musical patterns. Though this project claims to honor the intentions of a video-maker it forgets about the intention of the music composer and remains therefor incomplete from a rhythmical perspective.

3.2.4. Mining Association Patterns between Music and Video Clips in Professional MTV

The last paper I want to introduce here in detail is "*Mining Association Patterns between Music and Video Clips in Professional MTV*" by Chao Liao, Patricia P. Wang and Yiming Zhang published in 2009.

This is another innovative paper which bases their automation theory on the so called "dual-wing harmonium model". "*Provided with a raw video and certain professional MTV as template, we generate a new MTV by efficiently inferring the most related video clip for every music clip based on the trained model*"[52] Mentioned in the introduction of this paper the team had two very interesting guideline for the project: 1) "*According to the study in*

[51] Yoon, Lee, Byun, 2009, p. 2

[52] Liao, Wang, Zhang, 2009, p. 1

Visual Story, the film industry makes their products more intense and interesting by contrasting video components like color, movement and rhythm, etc"[53] and 2) "*Another study in further concerns the repetitive sections as prelude, interlude and coda in MTV, which can help to make the results more reasonable on semantic aspects"*[53]

We can see that it is one of the first automation researches dealing with the matter of rhythmic knowledge from professional use. It is further claimed that with this method video and music are strongly coupled (actually treated as pairs in the training phase) and matched based on how good they fit discovered association patterns, which are learned from a large dataset of professional MTV. These patterns are supposed to partially reflect professional MTV editor's skills and should both help in the content selection and composition problem in automatic MTV generation.

The Working Process in Detail:

First the professional MTV sample is divided into small clips. From these clips audio and video features are extracted to train the dual-wing harmonium model, which is therefor capable of representing semantic aspects. Vector V denotes the features (extracted from the color and structure tensor histogram) of the video clip. Vector U denotes the features (extracted from time and spectral domain) of the music clip.

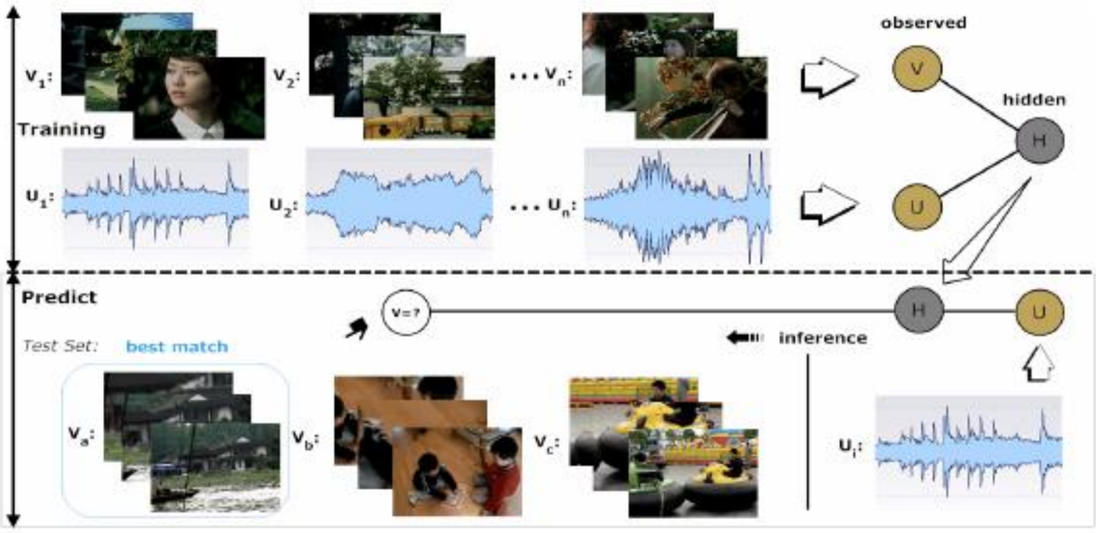

Figure 5 Framework of the Dual-Wing Harmonium Model

[53] Liao, Wang, Zhang, 2009, p. 1

"*The rules following which video and music clips are associated together are represented by the hidden node in the H space*"[54]. By clustering these points, several groups of similar clips can be figured out. "*These groups are considered as association patterns, and utilized to guide automatic MTV generation*."[54] In the predicting process the model will infer the most related video clip in raw video for a given music clip.

Results:

According to the paper extended experiments have been exercised with this model. Liao et al. collected over 100 manually generated MTVs in different categories from both male and female singers, both Chinese and English and both single and band group. These MTVs were segmented into video and music clips with the same time length, instead of dividing the segments by shots, which helped to avoid length balance problems when video and music differed from each other. From each video clip they extracted a 512 dimensional color histogram and a 17 dimensional structure tensor histogram. Here is a short explanation of what features are extracted by it.

"*The color histogram represents the color distribution as well as color richness in the video clip. The structure tensor histogram represents the motion intensity and direction in the video clip. Both of them are important measure for visual highlights. For each music clip, we extract its zero crossing rates in time domain, and its centroid, spread and the first 10 MFCC coefficients in spectral domain. We also extract the power, flux and low energy ratio in spectral window as its features. All these features are normalized to guarantee they have unit-variance*"[55]

After the training the video clips are mapped to H space, to get a low dimensional representation of the input data. Then data is clustered to association patterns. This was performed for both a single album and on whole data set. It was found more interesting association patterns were detected in a single album than in the whole set, which proofs that products from the same editor follow the trained editing rules better. One of the general observations here was that the semantic level for the discovered pattern depended on the features chosen. "*The audio features chosen in our experiments have the power to discriminate different people's voice, because they are also widely used in speech recognition*."[55]

[54] Liao, Wang, Zhang, 2009, p. 3

[55] Liao, Wang, Zhang, 2009, p. 8

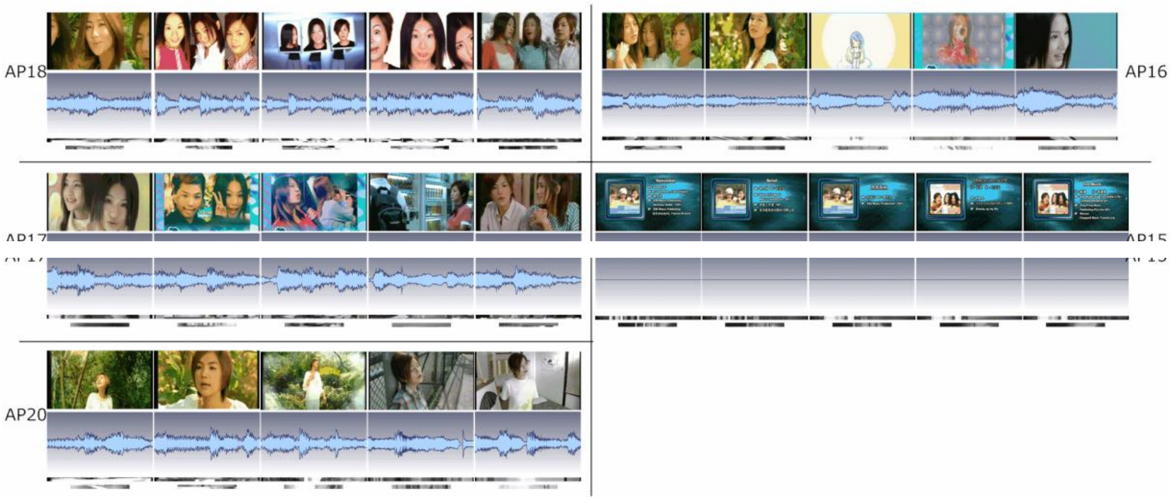

Figure 6: Association Patterns extracted from 10 professional MTVs of the Band S.H.E.

AP18: three girls' faces are shown when all of them are singing
AP17: only two girls are singing, and only two faces are shown
AP20 & AP16: both capture the solo scenes where only one girl is singing.

While in some patterns audio features dominated the selection, video features appeared to be stronger in others. By training the model with the whole data set even more complicated patterns occurred:

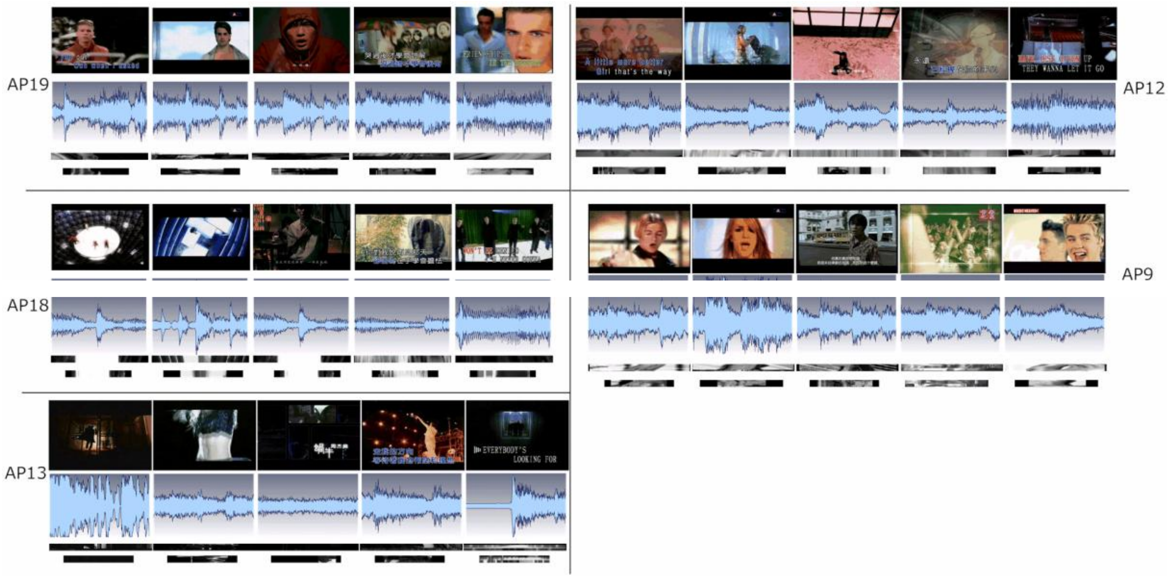

Figure 7: Association Patterns when model was trained with the whole data set

"*Since these MTVs are produced by various editors, the data points in the H space* *can be even harder to group together, and the discovered patterns should be considered as common rules followed by all editors*."[56]

AP19: Music highlights matched small motion of singer's head

AP9: also captured music highlight but with richer motion

AP18: camera barely moved and the music is not intense

AP13: loud voice well matches with dark decorating scenes.

AP12: is one pattern representing shot change

Remarkable patterns, though the paper states that since neither video nor audio features are dominant in these patterns, the underlying rules are somehow difficult to understand.

Except the function testing Liao et al. also tried to test the results objectively. Therefor they compared the similarity to the input professional MTV sample. Furthermore they applied a subjective user study to evaluate the effect on the audience.

To measure the similarity between an auto-generated result and the original MTV two aspects were considered: Exact match and Histogram divergence. "*Exact match is the total count that two sequences have same pattern labels in all positions. Histogram divergence is the K-L divergence between the histograms of two sequences… The* smaller the Histogram divergence is, the more similar two sequences are."57

The scores show that the results were more similar to original MTV than to the commercial automation program MUVEE in both Exact match and Histogram divergence. Recall the second guide line in MTV generation, the sequence of the original MTV contains the information for repetitive sections such as prelude, interlude and coda. In this sense, the method complied with the second guide line better than MUVEE.

Interesting results were also achieved on the subjective evaluation. While the users thought that the results were better matching to music than those of MUVEE the users felt that the generated MTVs were less exciting in their appearance. Liao et al explained this the following:

"*According to the users, the number of scenes in our result seems not as many as in MUVEE's. This actually reflects our method's limitation that we cannot guarantee shot consistency or scene loss in the result. We can further improve the final quality by collaborating with video highlight selection algorithms*"[57]

[[56]] Liao, Wang, Zhang, 2009, p. 9

[[57]] Liao, Wang, Zhang, 2009, p. 10

Conclusion & Rhythmic Evaluation

In comparison to the previous papers this research approach is completely different. While the others tried to imitate the human approach by adding static editing rules this project used real editing examples to train models and derive rules from it. A compelling and promising approach not only because this seems the most natural way (as we human beings also learn from existing knowledge) but because it is the most flexible approach to adjust to the individual needs of the high variety of music video genres. As we could see the association pattern for music and video clips in a MTV depend on many factors, such as its editor's taste or the music style. However, depending on which features were chosen, the discovered pattern can sometimes be hard to understand, when no feature in the pattern is dominant. And because video clips were divided with fixed time length, shot consistency cannot be guaranteed. What is so impressive on working with artificial intelligence is that rhythm as a complex topic does not have to be understood completely for being imitated. In other words, associative patterns do not have to be examined on how they work - we just have to know if they work.

So how can this scientific approach be improved. Liao et al. suggest the following in their conclusion: "*Our future work will be introducing the third modal data: lyrics into our framework, and combining our method with other highlight selection algorithms to further improve the result quality*."[58]

3.2.5. Web-based Approaches for Music Video Automation

By searching the internet I found several papers which are based web searching for finding music video source material. "*Automated Music Video Generation Using Web Image Resource"* by Rui Cai, Lei Zhang, Feng Jing, Wei Lai, and Wei-Ying Ma and *"Dancerproducer: An Automatic Mashup Music Video Generation System By Reusing Dance Video Clips On The Web"* by Tomoyasu Nakano, Sora Murofushi, Masataka Goto and Shigeo Morishima are only two examples of this approach. Though they are very interesting it would go beyond the scope of this paper to treat them into detail. Generally speaking it can be said that these approaches have one big advantage. They use already tagged video material which means that semantics can additionally be detected over text.

[58] Liao, Wang, Zhang, 2009, p. 11

Another big opportunity which is not helpful though for the generation of professional music videos therefor my question of research due to the fact the music videos are mostly generated out of unused material.

3.2.6. Existing Music Video Automation Software

As already mentioned several times, we can already find some automation software in the commercial sector. Generally purposed for home video editing, these algorithms mostly remain unpublished and therefore cannot be treated in my research. Examples for some of the most popular ones are: “MUVEE”, “First Cuts” or “Magisto”.

4. Conclusion

As we understand the importance of rhythm for our lives and especially for audiovisual media we can see that rhythmic approaches are essential to any kind of editing mechanism. Especially for music videos we can find associative patterns between audio and video that correspond in a rhythmic way and are essential for any professional and success-seeking appearance. During my studies I discovered that automation research has developed immensely in the past 10 years. Basic attempts in early 2000 have been rapidly outdated by more and more efficient approaches to imitate human editing. But while most of them tried to apply static editing rules out of empirical knowledge the newest and most promising ones deal with artificial intelligence to generate more and more accurate results in analogy to the human editing process. Still until now algorithms fail to achieve what manually edited music videos exercise every day – to thrill their audience. Reasons for this can of course be found in the vast variety of appearance and the indefinable nature of genre music video but some mistakes, in my opinion, can also be found in the incoherent rhythmic approach of each research project. While some have pretty advanced measures to capture the rhythmic information given by video and music sources, nevertheless all, examined in this research, fail to provide a complete set of rhythmic features discovered by empirical knowledge. Furthermore, of course, every automatic process has to fight the problem of mathematical consistency, which means that actions are always derived from certain rules, while the human editing process is everything else but consistent. In spite of all this, I believe there is a promising future for intelligent editing as seen in the last paper. If algorithms are improved by enhancing quality and quantity of the feature I believe there is even place for some specialized products in the professional market. I'm positive that a semi-automatic approach will soon appear to support editors in their process and that there will be an extended use of interactive editing programs. However from today's point of view a universal solution for automating edits which is comparable to the human approach is not likely accomplishable.

Bibliography

List of Literature

- Beebe, R. and Middelton, R., "*Medium Cool: Music Videos from Soundies to Cellphones",* Duke Univ Pr, 2007
- Bordwell, D. and Thompson, K. – "*Film Art: An Introduction*", 5th edition McGraw-Hill, 1997
- Cutietta, R. "*Using Rock Music Videos to Your Advantage*". Music Educators Journal 71 (6): 47–49, 1985
- Dancyger, K. – "*The Technique of film and Video Editing: Theory and Practice",* 5th edition, Focal Press, 2011
- Foote, J. ; Cooper, M. and Girgensohn, A. – "*Creating Music Videos using Automatic Media Analysis*", Published in: Proceedings of the tenth ACM international conference on Multimedia, 2002
- Girgensohn, A. ; Boreczky, J. ; Chiu, P. ; Doherty, J. ;Foote, J. ; Golovchinsky, G. ;Uchihashi, S. and Wilcox, L. – "*A Semi-automatic Approach to Home Video Editing",* Published in: Proceedings of the 13th annual ACM symposium on User interface software and technology, 2000
- Modell, A.H. – "*Imagination and the Meaningful Brain*", MIT Press, 2003
- Murch, W. – "*In The Blink of an Eye – A Perspective on Film Editing*", 2nd edition, Silman-James Pr, 2001
- Liao, C.; Wang, P.P. and Zhang, Y. – "*Mining Association Patterns between Music and Video Clips in Professional MTV",* Proceedings of the 15th

International Multimedia Modeling Conference on Advances in Multimedia Modeling, 2009

- Pearlman, K. - "*Cutting Rhythms - Shaping the Film Edit*", 1st edition, Focal Press, 2009
- Restak, R. – "*The New Brain: How the Modern Age Is Rewiring Your Mind*", Rodale, Emmaus, 2003
- Sosongko, J. – "*Automatic Generation of Effective Video Summeries*", Master Thesis, Queensland University of Technology, 2011
- Yoon, J. C.; Lee I.K. and Byun S. – Chapter 17 "*Automated Music Video Generation Using Multi-Level Feature-Based Segmentation"* of "*Handbook of Multimedia for Digital Entertainment and Arts*" Part 2, pp. 385-401, Springer , 2009

- *"The Compact Edition of the Oxford English Dictionary*. **II**.", Oxford University Press, 1971

List of Internet References

- BBC News, 07/20/2010, „Mother's heartbeat 'synchronises with foetus'" http://www.bbc.co.uk/news/uk-scotland-north-east-orkney-shetland-10696611 [01/12/2012]
- "Nonprofit Video: How Rhythm Keeps Us Watching" by Cam Hayduk and Kat Kelly http://www.nten.org/blog/2011/10/10/nonprofit-video [01/12/2012]
- Ramachandran, V.S., "Mirror Neurons and imitation learning as the driving force behind "the great leap forward" in human evolution". Edge Foundation. http://www.edge.org/3rd_culture/ramachandran/ramachandran_p1.html [11/16/2006].

- "Music Video Production Tips" – About Video Editing http://www.aboutvideoediting.com/tutorials/music-video.shtml [01/19/2012]

Table of Figures:

Figure 1: Automatically editing video to synchronize with a shorter audio track 20

Figure 2: Automatic music video system block diagram 20

Figure 3: Flow chart of Automatic Music Video Generation 22

Figure 4: Overview Working Process 26

Figure 5 Framework of the Dual-Wing Harmonium Model 29

Figure 6: Association Patterns extracted from 10 professional MTVs of the Band S.H.E. 31

Figure 7: Association Patterns when model was trained with the whole data set 31