A Method for Subjective Analysis of Audiovisual Works

Marko Ciciliani
Institute of Electronic Music and Acoustics
University of Music and Performing Arts Graz
ciciliani@iem.at

ABSTRACT

This paper proposes a method for the analysis of audiovisual works.The method of semantic differential is used in order to develop a differentiated understanding of the analyzed audiovisual work. A total of 8 parameters are distinguished, which are grouped in three categories that refer to categories of mapping, semantics and atmosphere.

The semantic differential yields a graphic display of the values that have been selected for the individual parameters of activity. Therefore an intuitive comparison between analyzed works is possible.

An alternative display on a time-line is also proposed, which facilitate the visualization of changes of parameters over time.

1. INTRODUCTION

Approaches to the analysis of audiovisual works and multimedia are relatively scarce and most of them focus on objectively verifiable aspects of the relationship between media, such as e.g. synchrony of events, or semantic correspondences [1]. While these aspects are undoubtedly of relevance for the perception of audiovisual works, they are not sufficient to reflect on the aesthetic impact of an instance of multimedia.

As several authors have argued, it is precisely the rather ephemeral effects that constitute the most fascinating side of audiovisuality, e.g. when sound "makes us see the image differently, and then this new image makes us hear the sound differently [...]" [2] or when "some or all of the [...] attributes of the one [medium] become available as attributes of the other" [3].

By referring to audiovisual works as the target of this proposed method, I indicate that the focus lies on interdisciplinary works where the perceived balance of music and visual media is equal, as e.g. in examples of visual music or many other experimental works of multimedia. Works where sound primarily supports an visual medium – as in many instances of narrative film – are not the focus of this research.

This paper tries to fill the aforementioned methodological gap by describing a method for a subjective analysis of audiovisual works. The focus of the method is the qualitative assessment of a subjective experience. The focus therefore lies on developing understanding rather than collecting objective facts about a particular work. Although objectivity is not the goal, the graphic display of the parameters makes it possible to intuitively compare the results of different audiovisual works.

The method lends itself well to sharpen the awareness how a specific audiovisual work develops a certain effect. It is therefore specially suitable for artists and researchers who wish to better understand with which means a particular work achieves an aesthetic impression.

2. THE USED METHOD

The method builds on terms that have been used for the description of multimedia by Cook [3], Iwamiya [1] and Kyriakides [4].

In his book *Analysing Musical Multimedia* Cook introduces a so-called *similarity test* and a *difference test*, through which he tries to assess whether sound and image amplify each other, which falls in the category of *conformance*; whether they try to dominate over each other, which corresponds to *contrast*; or whether they project their attributes to the other medium, which he calls *complementation*. As valuable as these differentiations might be, they are not very helpful when analyzing audiovisual works. The reason is, that these categories are not only very broad, but that most successful instances of multimedia fall into the category of complementation, as Cook acknowledges himself.

In the chapter "Perceived congruence between auditory and visual elements in multimedia", Shin-Ichiro Iwamiya differentiates degrees of congruence in the categories referring to formal, semantic, and mood-related aspects. These categories already make more differentiations possible than Cook's. However, Iwamiya's approach to multimedia explicitly values only conformances between media and does not recognize any sort of divergence between media as an artistically relevant choice. This reveals Iwamiya's background which lies in psychology and not in artistic practice. Therefore his approach to analysis is not suitable for the analysis of artistic works of multimedia, at least not in unaltered form.

In his yet unpublished work *The Imagined Voice*, Kyriakides distinguishes two categories of parameters for the analysis of multimedia, called perceptual and semantic [5]. In the former category he differentiates temporal and spatial, in the latter linguistic and narrative convergence and divergence. While Kyriakides' approach is very sensitive to artistic applications of multimedia, in my opinion it ignores some other aspects, that are relevant for the analytic description of audiovisual works.

This paper introduces a method for the analysis of multimedia, that consist of a combination, variation and extension of the parameters used by the aforementioned three authors. In addition, it uses the design of a semantic differential, where the impact of the various connoted parameters of the observed work are differentiated and graphically marked. This design is used because it has proven to be appropriate to derive the attitude of a subject towards a given object [6]. By differentiating factors of *valence* (good-bad), *activity* (active-passive) and *potency* (strong-weak) a better understanding can be achieved of how a given instance of multimedia is experienced by a subject.

In the proposed method a total of 8 parameters of activity are grouped in three categories of potency: *parameters of mapping, parameters of semantics* and *parameters of atmosphere*. By indicating the degree of activity of each group, the relevance can be shown that the parameters represented by the group hold in the analyzed audiovisual work. The parameters and its categories are presented in Tab.1.

| Category of Mapping	Category of Semantics	Category of Atmosphere
Synchrony	Congruence of **content**	**Kinetic congruence**
Spacial congruence	Coherence of **idiom**	**Balance of salience/fidelity**
Congruence in terms of **material**		Correspondence of **tinting**

Table 1: Factors for the analysis

Eventually, the experienced overall coherence of the observed work can be indicated as a factor of valence. Graphically the method is arranged as displayed in Fig.1.

All values are marked graphically on an axis. With the parameters of *potency* and the axis for *valence*, the values are arranged around a midpoint which describes a neutral situation. In contrast, the axis for *activity* runs continuously from a minimum to a maximum value. The axis have markings at regular intervals for better orientation but do not show a numeric scaling, in order to underline the subjective and qualitative nature of the method.

Contrary to many applications of the semantic differential, where the examination starts with *valence*, then con-

Valence, Potency, Activity
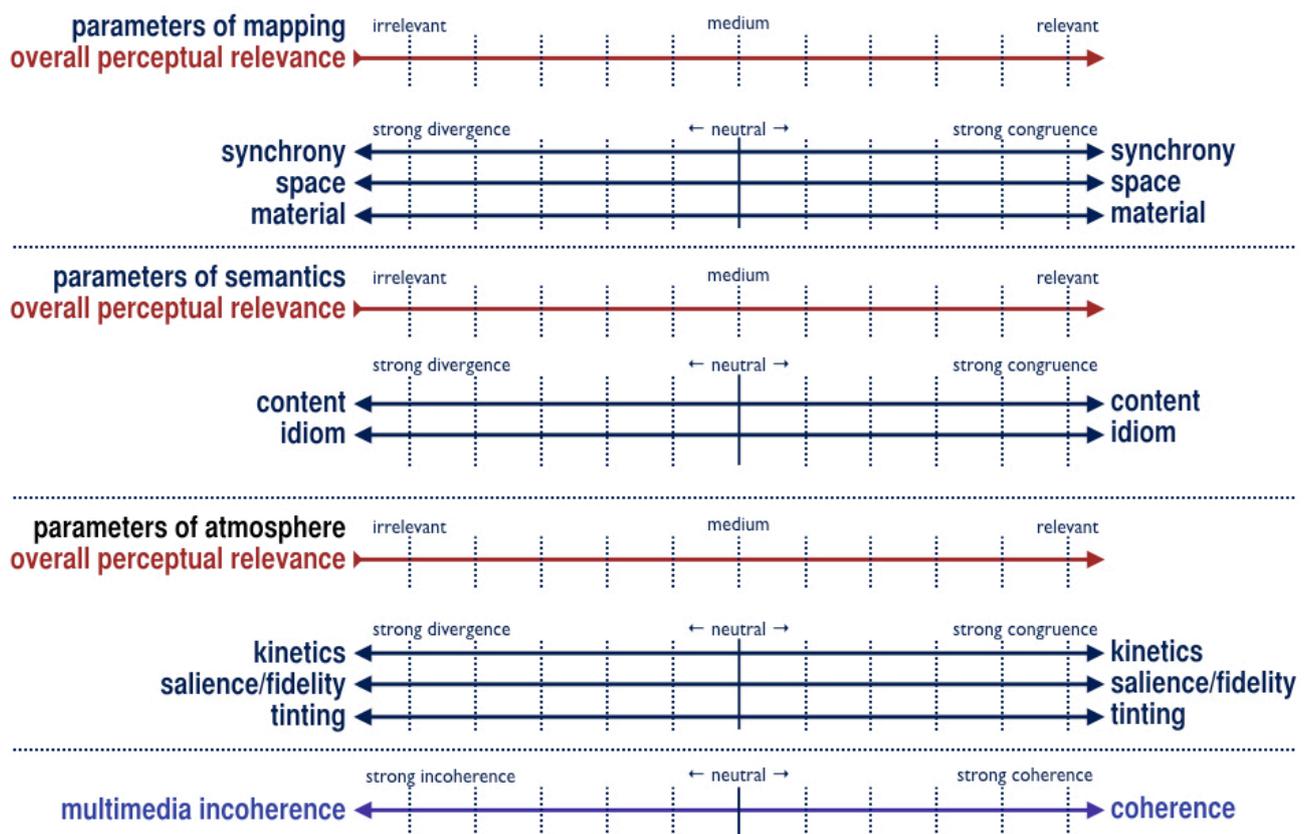
Figure 1: Graphical arrangement of the analysis factors

tinues with *potencies* and ends with *activities*, in the case of this analysis method it is essential to proceed backwards: start with the parameters that show the *activities* of a number of details, progress via *potency* in order to assess the aesthetic significance of a category in the given context, and end with the overall *valence* of the work. The reason for progressing in this particular order is that the categories of *potency* are subsuming the parameters of *activity*. Since the categories are therefore arranged hierarchically, an understanding of the details is necessary in order to estimate the magnitudes of the broader categories.

3. DISCUSSION OF THE CATEGORIES

Three main categories are distinguished in this method that address various aspects of the relationship between visual and auditory events.[1] All parameters belonging to a category are displayed on an axis where the midpoint expresses a neutral state, which means that this particular parameter does not apply or that it is not designed in a salient manner. If the observed media relate to each other congruently, this is indicated on the right half of the axis. Congruence can also be described as a mutually supportive or amplifying effect that the media have on each other. If the media are divergent, this is indicated on the left half.[2] Divergence is at hand when there is some sort of friction between the two layers.

3.1 Category of Mapping

Three parameters are subsumed in this category that refer to congruencies or divergences between media in terms of timing, spatial arrangement or size. These parameters are crucial for creating an effect of audiovisual source-bonding.

3.1.1 Divergence/Congruence in terms of synchrony
If it is desired to create a perceptual fusion where a visual event is associated with an aural one, it is necessary that they occur simultaneously. Michel Chion famously coined the term *synchresis* for this "spontaneous and irresistible weld produced between a particular auditory phenomenon and visual phenomenon when they occur at the same time" [7]. Accordingly, the first parameter in this group describes congruencies or divergences in terms of synchrony. Here it is indicated whether auditory and visual phenomena are predominantly synchronized. While in general applications of sound design, synchronization is usually applied to create effects of a common cause – as for example in Norman McLaren's *Dots* (1940) –, synchronization is often used as a general design principle to create cohesion between music and image, as for example in Granular Synthesis' *Modell 5*.

Prevalence of synchronous arrangements would therefore be indicated on the right half of the axis. If both sound and vision would show clearly pronounced rhythmic structures that were not lined up in any way, this would be described as divergence. The neutral mid-point would be chosen if no discernible rhythmic relationship would exist between the two, as for example if one would be static and the other would go through rapid modulations.

3.1.2 Divergence/Congruence in terms of space
The second parameter of this group refers to space and indicates if the spatial arrangement of sounds and elements in the image correspond spatially. An example for this can be found in Ryoichi Kurokawa's live performances with two video screens, where the horizontal motion often corresponds with the stereo panning of the sound. This category can also be suitable for indicating correspondences between visual space and pitch space, e.g. when higher frequencies are aligned with events in the upper region of a screen, and low frequencies in the lower region.

3.1.3 Divergence/Congruence in terms of material
The third parameter indicates if sonic and visual phenomena correspond in terms of material. A typical example would be that a larger object is associated with a lower or louder sound than a small object, or that surface characteristics of a depicted object are sonically emulated.

3.2 Category of Semantics

Contrary to the other categories that contain three parameters, respectively, this one only contains two: divergence/congruence in terms of content, and of idiom. This category refers to references that are pointing beyond the work itself. This can pertain to specific meanings or even narratives that are conveyed through the work, or more abstract references to other contexts. Indexical or intertextual references can take on a myriad of different forms and this category does not presume to reflect that diversity. The two parameters are indicating, though, whether divergences or congruences are formed on the level of content or idiom:

3.2.1 Divergence/Congruence in terms of content
This parameter indicates whether an audiovisual work exhibits a profiled relationship between media of a semantic nature. Here semantic refers to any meaning that is superordinate to what is evident as a visual or aural phenomenon. These can be a concrete meanings – as in an narrative context, a symbolic reference, or emergent meaning. With emergent meaning I am referring to new meanings that arise through a particular constellation or treatment of audiovisual phenomena. An example would be the scene from Martin Arnold's *Alone. Life wastes Andy Hardy* (1998) with a young boy and his mother, which takes on erotic qualities with an "incestal subtext" [8] through the looping of sound and image.
An interesting example for divergent semantics can be found in Robert Cahen's *Karine* (1976). Here film footages and photos of his daughter's childhood are arranged as a filmic photo album. While both sound and

---

[1] It would be possible to apply this method also for other combinations of media, as for example sound and touch. The focus in this paper lies, however, in the application on audiovisual artistic works.

[2] At this point I would like to emphasize that none of these settings expresses any judgement. Neither is congruence better than divergence or vice versa, nor is a strongly pronounced parameter any better than a neutral one.

image work with evocations of past time, the film has a nostalgic and poetic quality, while the music mainly focusses on crackling sounds that in my opinion display a more destructive and erosive passing of time and altogether creates uncanny associations that contrast the sentimentality of the film.

3.2.2 Divergence/Congruence in terms of idiom

This parameter expresses whether there is congruence or divergence between the idioms or styles that each medium uses. A typical example for this would be a baroque opera with a modern staging where music and stage design use completely different stylistic design elements. An example of idiomatic divergence from the repertoire of visual music is Lillian Schwartz's *Googleplex* (1972), where digitally created, strongly pixellated images are combined with non-western percussive music of gestural kind.

3.3 Category of Atmosphere

Three parameters of atmosphere are describing characteristics of relationships between media, that are rather evasive and that can not easily be attributed to a single concrete principle. The term *atmosphere* is used in the meaning as described in Gernot Böhme's aesthetics of atmosphere [9]. In his book *Atmosphäre* he describes atmosphere as "something vague, diffuse, however not vague in reference to what it is, namely its character."[3] [10]. Böhme describes atmosphere as something which is in between objects and which can not be attributed to any single element. Instead, it is poured out between subject and object and despite of its ephemerals nature, it is very specific. In the application to multimedia the three parameters of atmosphere attempt to give a better understanding of the character of the work and how it is evoked. The parameters refer to kinetics, salience and fidelity, and tinting.

3.3.1 Divergence/Congruence in terms of kinetics

This parameter describes the temporal relationship between media. As it refers to temporality it has similarity with the parameter of synchrony (3.1.1). However, instead of describing whether or not elements from different media are temporally aligned, kinetics refer to a more general sense of motion, pacing and speed. An example for a pronounced congruence in terms of kinetics can be found in Donn Alan Pennebaker's *Daybreak Express* (1953) where sound and image are never exactly aligned, but where they are nevertheless strongly matched (also – by the way – indexically in terms of semantics!). An example for a strong divergence in terms of kinetics is practically any of Phill Niblocks films from his DVD release *The movement of people working*. While the music is usually based on elongated and drone-like textures, the images show people at work and therefore in constant motion. Here, the general sense of kinetics is therefore divergent. As I will show further below, Niblocks works nevertheless display strong congruence in terms of tinting.

3.3.2 Divergence/Congruence in terms of salience/fidelity[4]

This parameter describes the balance between different media and whether one strongly dominates over the other (salience), or whether one is designed in much greater detail than the other (fidelity). In audiovisual works, where the visual and musical layer are usually treated as equal in importance, this parameter therefore tends to be on the congruent side. There are however examples, where divergence can be can be detected, An example for divergence in fidelity is Mary Ellen Bute's *Synchromy No.4* (1938), where Bach's music is mapped to geometric shapes, that clearly show a lower level of detail than the music.

3.3.3 Divergence/Congruence in terms of tinting

This parameter addresses perhaps the least tangible aspects of a work, namely the general moods that the media express – here expressed as tinting. An example of strong congruence in terms of tinting would be Tony Conrad's *Flicker*, where sound and image are quite unrelated in terms of mapping, yet both express a very raw quality that works very coherently as a combination. Also Phill Niblock's aforementioned films display strong divergence between sound and image not only in terms of kinetics but also content. A strong semantic divergence is formed by the combination of the display of people working in apparently non-western countries, combined with an almost meditative music played on western instruments. Yet, in my opinion Niblock's films correspond very fascinatingly in their general mood, as both are characterized by repetitive phenomena that are non developmental. In two entirely different ways, film and music are characterized by situations that are rather static, although on a level of detail they also display constant variation. An example of divergence in tinting would be William Raban's *Broadwalk,* (1972) where sound and image individually evoke different atmospheres, which – in my opinion – also does not fuse or converge when they are combined, although this particular combination works strongly.

4. TIME SCALE OF THE PARAMETERS

The different parameters are addressing different features of an audiovisual work. However, the boundaries between them are often fluid. Therefore they should not be interpreted rigidly, but rather as offering different perspectives on how a work can be observed. Each of the three categories also observes occurrences on a different timescale. The category of mapping is concerned with relationships between phenomena that are taking place on a short time scale. Congruences or divergences of synchrony, the use of space or sizing are limited to short gestural or even momentary events. The parameters of atmosphere, however, can only be grasped while observing longer time intervals. Because of the often ephemeral nature of atmospheres they can rarely be identified in very short time intervals. However, the parameters of indices

---

[3] My translation

[4] I took the term *fidelity* from Broadwell and Thompson's use in film theory (Broadwell/Thompson 1993:304)

can manifest in very different time spans. The use of clearly identifiable symbols communicates instantaneously, while other semantic aspects can evolve over long periods of time, as e.g. the ongoing descending melodies in section "I" of Fausto Romitelli's *An Index of Metals*, which semantically corresponds to the sung text, which describes the process of drowning.

5. FACTORS OF POTENCY

Each of the three groups of parameters forms a factor of *potency*. Factors of potency express the relevance that a group of parameters holds in a work. A work might e.g. exhibit strong correspondences in its mappings in terms of synchrony and spatial motion, while this can still remain a secondary aspect of what characterizes it as a whole. Therefore there is no direct correlation between the settings of the parameters and the actual potency that a group of parameters holds in a work.[5]

Choosing the magnitudes of the potencies is the most important step in the analytic process. It forms an interpretation of the relevance of the various aspects that have been observed. Therefore it is mandatory to first investigate the individual parameters, before their overall importance in

the given context is indicated.

The factors of potency thus reflect the aesthetic significance that a particular feature or group of features hold in the analyzed work.

6. VALENCE

Finally, the analysis concludes with determining the overall valence of the observed work. This is expressed as the degree of coherence that a work displays. Again, I would like to emphasize that this is no judgement of the work. An incoherently designed audiovisual composition can be artistically as strong as a coherent one. Although this might be debatable, I would argue that Michael Snow's work *Wavelength* (1967) is very incoherent in how the visual and sonic phenomena were combined, yet, this is exactly one of the strengths of the film.

When a single passage in an audiovisual work is observed, the indication of valence often seems awkward and arbitrary. However, when comparing different audio visual works, or the development of a single work over time, this aspect can become very meaningful. I'll return to this in the next subchapter.
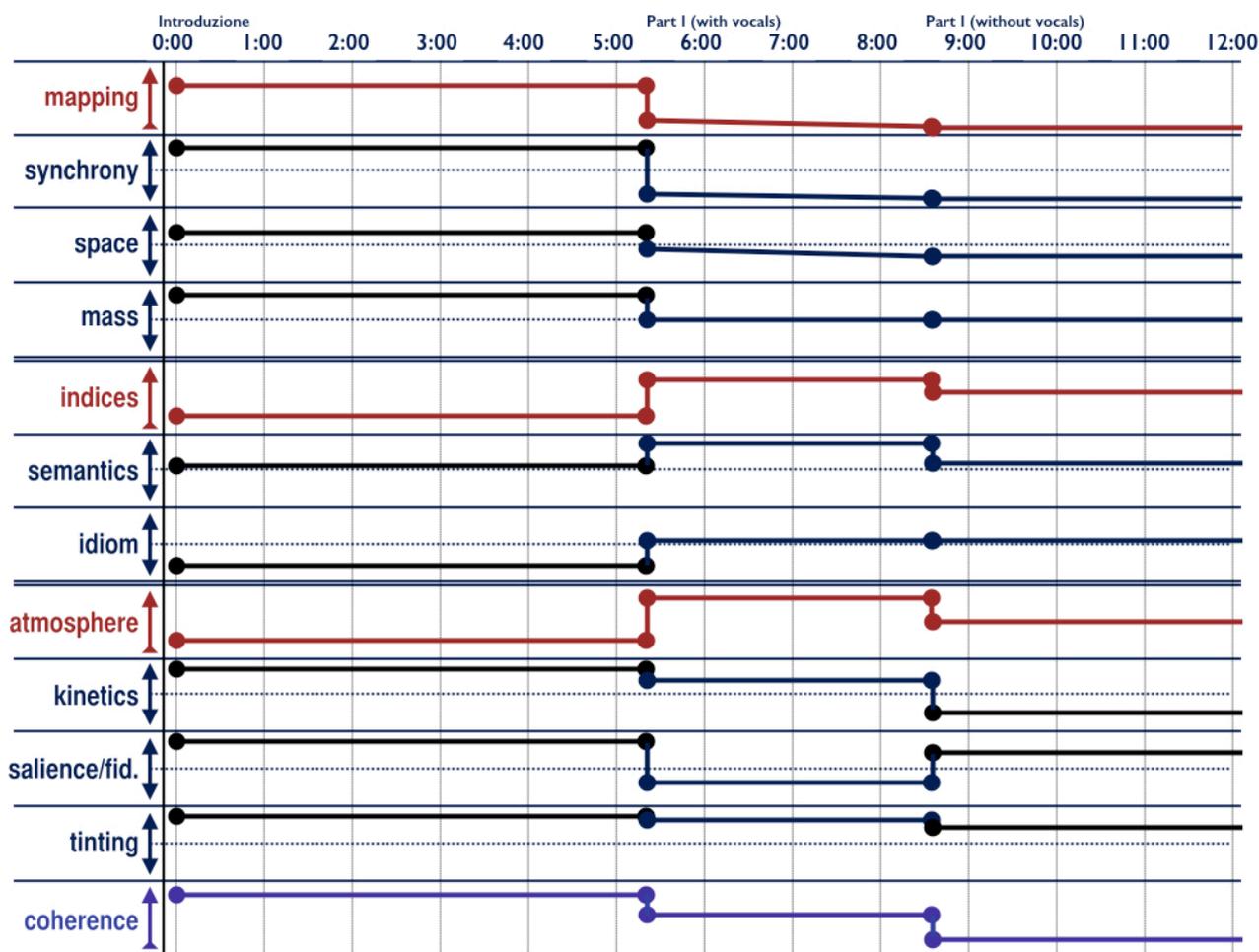

Figure 2: Analysis of the first two sections of Fausto Romitelli's video opera *An Index of Metals*

---

[5] This is true except when all parameters of a group remain in the midpoint of the axis, meaning that they neither exhibit strong congruence, nor divergence. In such a case it is very unlikely that this group will have strong potency within the work.

7. CHANGES OVER TIME

The form displayed in Fig.1 describes and displays a specific relationship between the auditory and the visual layer in an audiovisual work. Since many works alter the relationship and the detailed constituency of its parts during the course of a piece, it is necessary to display changes of parameters, potencies and valences over time. The graphic display used so far is not suitable for this.

Alternatively, the values of the different aspects of the analysis can be arranged vertically, while their corresponding values are displayed over a horizontal time axis.

This form of display is similar to the visualization of automation tracks in e.g. digital audio wavestations. *Figure 2* shows as an example the development of the parameters and categories of an analysis of the first 12 minutes of Fausto Romitelli's *An Index of Metals* (2003).

8. SUMMARY

The described method is not primarily intended to assess objective facts of a work of multimedia. Rather its goal is to help understand how a work develops a certain effect on the analyzing subject. The method is therefore primarily intended as a method for subjective analysis.

The method is based on a combination, variation and extension of parameters and categories used by Cook, Iwamiya and Kyriakides. Apart from introducing some new terms and sharpening the meaning of the others, I believe that a more detailed analysis is possible with the proposed model. The effectiveness is increased furthermore by using the design of the semantic differential, which offers a progressive assessment of the qualities of a given work, rather than a piling up of attributes. The three categories of valence with their total of 8 parameters give good points of orientation to develop an understanding of the most relevant – in the sense of character or expression forming – features of a work.

The indicator for *valence* may be less significant when a singular moment of a work is observed. However, it can especially become meaningful with longer analyses where changes over time are displayed as in Fig.2.

9. REFERENCES

[1] Iwamiya, Shin-Ichiro: "Perceived congruence between auditory and visual elements in multimedia" in Tan, Siu-Lan et al.: *The Psychology of Music in Multimedia*, Oxford University Press, 2013, p.141

[2] Murch, Walter: "Introduction" in: Chion, Michel: *Audio-Vision*, University Press Group Ltd, 1994, p. xxii

[3] Cook, Nicholas: *Analysing Musical Multimedia*, Oxford University Press, 1998, p.69

[4] Kyriakides, Yannis: *The Narrative Voice*, unpublished, 2015

[5] Kyriakides, Yannis: *The Narrative Voice*, unpublished, 2015, p. 57

[6] Osgood, C. E., May, W. H., and Miron, M. S.:*Cross-Cultural Universals of Affective Meaning*. Urbana, IL: University of Illinois Press, 1975

[7] Chion, Michel: *Audio-Vision*, University Press Group Ltd, 1994, p.63

[8] Howard, Ed: "Films I love #13" www.seul-le-cinema.blogspot.co.at [March 24, 2016]

[9] Böhme, Gernot: *Atmosphäre*, Frankfurt: Suhrkamp, 2013

[10] Böhme, Gernot: *Atmosphäre*, Frankfurt: Suhrkamp, 2013, p. 22

[11] Broadwell, David, Thompson, Kristin: *Film art: an Introduction*, McGraw-Hill, New York, 1993

Audiovisual examples:

Arnold, Martin: "Alone. Life wastes Andy Hardy", 1998, https://www.youtube.com/watch?v=LgqH3PK6-3Q [May 17, 2016]

Bute, Mary Ellen: "Synchromy No.4: Escape", 1938, https://www.youtube.com/watch?v=YR-mu-GcClls [May 17, 2016]

Cahen, Robert: "Karine", 1976, in: *Robert Cahen Films + Videos 1973-2007*, Geneve: Ecart, 2010

Conrad, Tony: "Flicker", 1966, in: *Dreaminimalist*, Paris: Re:Voir Video, 2011

Granular Synthesis: "Modell 5", 1994, in: *Immersive Works 1992-2002*, Karlsruhe: ZKM, 2004

Kurokawa, Ryoichi:"syn_", 2011, https://vimeo.com/61347519 [May 17, 2016]

McLaren, Norman: "Dots", 1940 https://www.youtube.com/watch?v=E3-vsKwQ0Cg [May 17, 2016]

Niblock, Phil: *The Movement of People Working*, 1975-1983, Preston: Extreme, 2003

Pennebaker, Don Alan: "Daybreak Express", 1953, https://www.youtube.com/watch?v=gDqBoUYpMq4 [May 17, 2016]

Raban, William: "Broadwalk", 1972, https://www.youtube.com/watch?v=cH9FCHILxQs [July 23, 2017]

Romitelli, Fausto: "An Index of Metals", 2003, Brussels: Cypres Records, 2006

Schwartz, Lillian: "Googleplex", 1972, http://lillian.com/1972-googolplex-5-12-min/ [May 17, 2016]

Snow, Michael: "Wavelength", 1967, https://www.youtube.com/watch?v=aBOzOVLxbCE [May 17, 2016]