The Relation between Music and Moving Images: A Cross-Paradigm Study

Paper for the workshop at the Audio-visuality conference, May 29, 2011

Anders Bonde, Ph.D., Associate Professor, Department of Communication and Psychology, Aalborg University

Abstract

Audio-visual media products, combining semiotic resources such as moving images and music, represent powerful semiotic artefacts, given that audiences' perception and reception bring forth affective meanings and connotations that transcend or deviate from the expressive and semantic/pragmatic meaning potentials of the contributing semiotic resources themselves. Such notion is applied to legal thought by theorists and analysts of film sound and music (e.g. Pauli, 1976; Chion, 1994; Cook, 1998) and social semioticians (e.g. Burn & Parker, 2003; Hull & Nelson, 2005), and it has been a subject of quantitative empirical research in experimental psychology (e.g. Marshall & Cohen; Libscomb & Kendall, 1994; Ellis & Simons, 2005). However, these varied disciplinary areas are traditionally interwoven with different philosophical paradigms and methodologies, and therefore they appear largely mutual alienated. With the belief that further knowledge can be obtained by paradigm combination and method triangulation, I conducted (in collaboration with Nicolai J. Graakjær) a study of audio-visual interaction, using various kinds of data. A video containing two alternative musical underscores as well as the visuals and music alone were each analysed for syntax, semantics and pragmatics, and evaluated by high-school students, writing down their associations freely (cf. Tagg, 1989), and filling out multiple-choice tests concerning *emotion* and *brand image* (cf. Hung, 2001). The results indicate a complicated picture where the inferred meanings of the audio-visual whole correspond to the contributing parts according to some variables, while being dissimilar in other aspects. After presenting the design of the study and its results, I will discuss the study with regard to methodological and epistemological issues, and as a conclusion, I will present some ideas about how the concept of 'emergence' might relate to audio-visual meaning and how to incorporate this in empirical research.

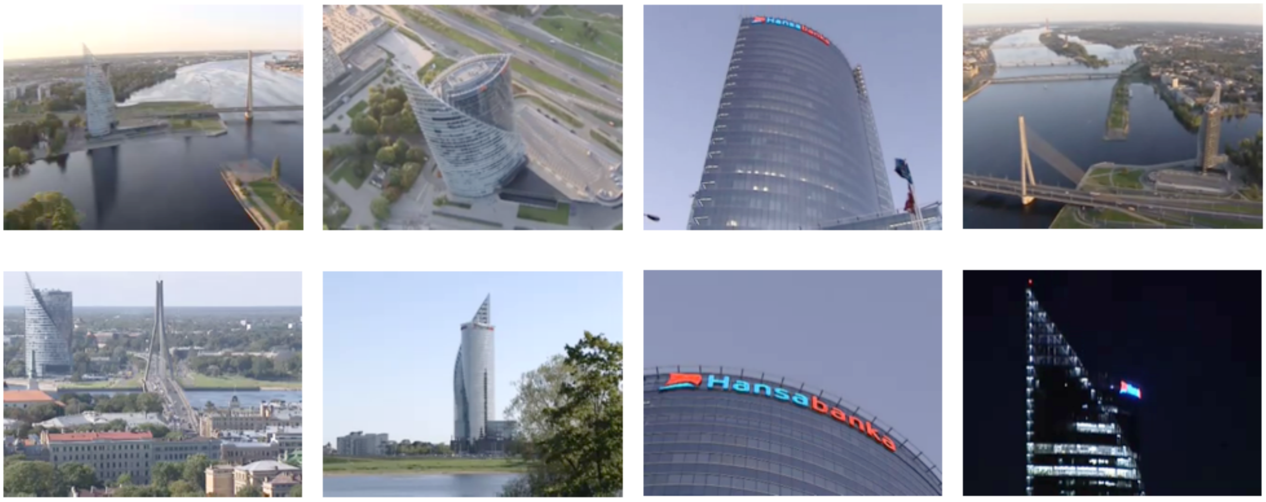
Introduction

It is a common notion among researchers in the disciplines of film music or film sound (e.g. Pauli, 1976; Chion, 1994; Cook, 1998), social semiotics (e.g. Burn & Parker, 2003; Hull & Nelson, 2005), and experimental psychology (e.g. Marshall & Cohen; Libscomb & Kendall, 1994; Ellis & Simons, 2005) that the perception and reception of audio-visual media products, combining different semiotic resources, bring forth coherent meaningful and aesthetic multimodal wholes that transcend or deviate from the expressive and semantic/pragmatic meaning potentials of the semiotic resources themselves. As such, the present research project contributes to a general theoretical understanding of the interaction between music and moving images by suggesting how to scrutinize or validate this relationship empirically. Departing from the theory of music in moving images as 'paraphrasing', 'polarizing', or 'contrapuntal' (Pauli, 1976, 104), I will present a recent study (in collaboration with Nicolai J. Graakjær) of audio-visual interaction, in which we have sought to avoid the typical 'hierarchical' understanding of music as a secondary semiotic device 'doing something to the images', the primary source, rather than the other way around (Cook, 1998). In this study we have combined different epistemologies and methodologies, integrating qualitative and quantitative approaches. For one thing we analyzed syntactical structure and semantic, connotative meaning potentials in two audio-visual productions, with the same images but different musical soundtracks; and for another we conducted self-reporting tests, in which the same productions were used as stimuli and evaluated by high-school students.

Generating stimuli

The source of the testing material is a 31 sec. video which was downloaded from the video-sharing website YouTube.[1] The video features the Saules Akmens building in Riga (headquarter of the Hansabank, now Swedbank Central Office, Latvia), and consists of a sequence of eight shots (including zoom effects), showing the building from different angles and perspectives (cf. Fig. 1). It was shot from air balloon by an amateur photographer and made as part of a documentary about wooden architecture of Riga.[2] He offered the video for the banking company like a promotional video, and in some sense it appears like an advertising film (because of the threefold depiction of the logo combined with the short duration of the video). However, the company never used it.

Figure 1



[1] http://www.youtube.com/watch?v=v9kC0qnDry0&feature=related

[2] Cf. e-mail correspondence, August 6 2010.

The choice of the video as suitable for use as testing material has several motivations. To begin with, the video contains no dialogue, voice over or any kind of real sounds: only music, which made it possible to insert alternative musical soundtracks and generate different audio-visual versions without losing any signifying elements in the diegetic framework. Furthermore, there are no verbal messages except for the Hansabank logo, and the video portrays no human faces or emotions. In fact, no people are depicted at all. Accordingly, the syntagmatic structure of the video is purely descriptive (Metz, 1974). We have recognized this as an advantage: With practically only moving images and music at work, and the former lacking emotional content, it seemed easier to gain knowledge about what is measured, and to trace the source of the inferred impressions and associative meanings. Secondly, the film depicts a foreign (not Danish) banking company, which is probably unknown to the majority of the Danish population. That made another advantage, seeing that any preliminary knowledge and viewpoints about the banking company was almost non-existing.

Two undergraduate students in Popular Music & Sound Production at Aalborg University have each independently manipulated the video, creating alternative musical soundtracks. As working source they were given only the moving images (without sound), and they were told to produce a musical soundtrack as if it was a TV commercial. On this background we expected two productions that would appear 'natural' despite the 'amateurness' of the video. The two versions include one with a new composed soundtrack without lyrics (the 'finger-snapping version'), and another containing a mix from the Lennon/McCartney song 'We Can Work It Out' (the 'The Beatles version'). They differ significantly in terms of 'codedness' (Middleton, 1990, 173f): Whereas the latter can be described as 'overcoded' because of the lyrics and the popular-song status, the former might, conversely, be categorized as 'undercoded' as a result of the music's 'unknownness' and the absence of lyrics. Consequently, we expected that the different levels of 'codedness' to influence the perception and reception of the audio-visual whole, and that the two versions would appear paraphrasing or polarizing, and contrapuntal respectively (Pauli, op.cit.).

Respondents, testing conditions and procedure

145 high-school students in 7 different classes (divided among 4 schools in Northern Jutland) participated in the tests, evaluating either one of the audio-visual versions, or one of the musical soundtracks alone, or the moving images alone. For practical reasons, the tests took place in classrooms equipped with media projectors. In each class the students were presented with only one stimulus, and they were told either to focus on the *film* or the *music*. That made up 7 testing conditions:

1. The evaluation of the *film* in the audio-visual whole of the 'finger-snapping version'
2. The evaluation of the *music* in the audio-visual whole of the 'finger-snapping version'
3. The evaluation of the *film* in the audio-visual whole of the 'The Beatles version'

4. The evaluation of the *music* in the audio-visual whole of the 'The Beatles version'
5. The evaluation of the *film* in the moving images alone
6. The evaluation of the *music* in the 'finger-snapping soundtrack' alone
7. The evaluation of the *music* in the 'The Beatles soundtrack' alone

Apart from the conditional differences listed above the testing procedure was identical in every class. The rationale for differentiating between the test conditions is twofold: First of all, we wanted to compare the audio-visual results with the results from the control conditions (moving images and soundtracks alone) to see whether the 'whole' transcend or deviate from the parts. Secondly, by letting different respondents evaluate the *film* and the *music* independently in both audio-visual versions, we would be able to investigate the influence not only of the music but also the moving images, and by that prevent the 'hierarchical fallacy' as mentioned previously.

The testing procedure was divided into three steps, each involving presentation of stimulus succeeded by self-reported responses. At first the students were told to write down what came into their mind (for about 2 minutes). This step was inspired by the 'free-induction' procedure used by Tagg (1989), and our reason for choosing this methodology was similar; i.e. "each answer has greater cultural and symbolic significance" when "created with the music [or video] as sole stimulus and not with the aid of ready made alternatives" (ibid., 23). For the second and third steps we developed our own multiple-choice schemes, adopting Hung's (2001) categorizational distinction between, on the one hand, affective-emotional responses regarding the respondent's personal feelings, and on the other, semantic-associational responses regarding the perceived signal values (ibid., 43).[3] For every single variable, concerning emotion or signal value, there were four answer choices: 1) 'No, on the contrary'; 2) 'No'; 3) 'Yes, to some extend'; and 4) 'Yes, very much'.

Results

The responses from the 7 testing conditions were stored and analyzed in *Survey Xact*, in which the average values for each affective-emotional and semantic-associational variable were calculated, and then compared across the conditions.[4] As for the free-induction responses, we employed a qualitative content analysis, categorizing the impressions and

[3] The multiple-choice scheme, concerning affective-emotional responses (step 2), included 16 variables adjusted according to the widely used 'circumplex model' (Russel 1980) that describes the relation between high/low arousal and positive/negative valence. The shaping of the other multiple-choice scheme, concerning signal-value responses (step 3), included 27 variables, of which we have sought a balance between positive and negative statements as well as 'hedonistic' vs. 'ascetic' values.

[44] *Survey Xact* is an online survey software tool developed by Ramboll Management Consulting Denmark (http://www.surveyxact.dk/). The four answer choices were quantified as -1.00 ('No, on the contrary'), -0.80 ('No'), 0.80 ('Yes, to some extend'), and 1.00 ('Yes, very much'), using negative and positive numbers to reflect the overall binary structure of the multiple-choice scheme.

associations of the respondents in five 'brand personality dimensions': sincerity, excitement, competence, sophistication and ruggedness (Aaker, 1997). Due to the scope of the present paper, I shall not present all the results, but rather briefly concentrate on few points.

Moving images vs. music
To begin with, the responses of the control conditions vary considerably. This can be assured by comparing the average values of the 16 and 27 variables in the 'moving-images-alone' condition with the 'music-alone' conditions across the control conditions. The emotional and semantic difference is particular revealing when evaluating the mixed Lennon/McCartney song alone, seeing that almost all of the freely induced impressions and associations seem to be linked with sincerity and excitement, while the moving images apparently induce *competence* and excitement instead. Secondly, while both musical soundtracks generally evoke positive emotions, the moving images seem to evoke no feelings at all. As for the signal values, there is, interestingly, an inverse proportional relation between the results of some of the 'hedonistic' and 'ascetic' variables. Indeed, variables such as 'complacency', 'classiness', 'trendiness', and 'glamour' are rated relatively *high* in the 'moving-images-alone' condition and relatively *low* in the 'music-alone' condition, while it is the other way around in the case of variables such as 'down to earth', 'simplicity', 'sincerity', 'trustworthiness' and 'naivety'. Hence, there are reasons to believe that the music has an effect on how the moving images are perceived, but also that the moving images influence the perception of the music.

The whole vs. the parts
It seems that the music evokes a feeling of 'good mood' in both audio-visual versions, but it also contributes to 'satisfaction', 'enthusiasm', 'naivety' and less 'uncertainty' as well as more 'stimulation' and less 'boredom', especially in the fourth test condition where the respondents are asked to evaluate the music in the 'The Beatles version' (see the list above). However, generally the perceptual relation between the visual and musical parts alone and the audio-visual 'whole' is a complicated and unpredictable matter that cannot be fully illustrated by using the typology of Pauli (1976); the relations basically depend on which variables one considers. Nevertheless, what seem interesting are those cases where the results of the audio-visual conditions transcend significantly the results of the control conditions; i.e. when the 'whole' is *much more* or *much less* than the parts. Here, the 'finger-snapping version' includes more profound examples than the 'The Beatles version' because of the larger emotional and semantic difference between music and moving images, and the fact that the evaluation of the audio-visual 'whole' tends to be a 'middle-of-the-road' result. Actually, there is an inverse proportional relation between 'audio-visual transcendence' on the one side, and the evaluative difference between music and moving images on the other: the lesser the difference, the larger the degree of transcendence; the greater the difference, the lower the degree of transcendence. There are other traceable patterns regarding the 'whole' as transcending or deviating from the parts, though in this paper I shall only point out a general one. The 'finger-snapping version' evokes less 'surprise', 'wonder', 'confusion', 'dishonesty'

and ‘unreliability’; but also less ‘inventiveness’, more ‘unoriginality’ and more ‘normalcy’. One might perhaps explain this as a consequence of the fact that this version includes several synch points between the moving images and the music, and therefore it appears more credible and authentic than the ‘The Beatles version’, of which several respondents, conversely, have noticed a ‘mismatch’ between *architecture in the new millennium* (the Hansabank pictures) and *nostalgia of the sixties* (popular music of The Beatles).

Transcendence or emergence

On the question whether the audio-visual versions transcend or deviate from the contributing parts, there is no simple answers, since the perception of music and moving images as mutually ‘paraphrasing’, ‘polarizing’, or ‘contrapuntal’ (Pauli, 1976) depends on focal point (variable). In this regard it is particularly noticeable that there is a strict line between the self-reporting method used for the study, and the way the whole-parts relation can be interpreted. As a matter of fact, whereas the multiple-choice results sometimes indicate ‘audio-visual transcendence’, the free inductions ‘make room’ for interpreting the audio-visual versions as signifying *differently* compared with the music and moving images alone – at least from a principal point of view. The latter is closely related to the phenomenon of *emergence*; i.e. an irreducible and coherent whole with novel qualities such as non-predictive effects or meanings. Now, in the humanities the concept of emergence is used to characterize the nature of ‘multimediality’ (Cook, 1998) or ‘multimodality’ (Hull & Nelson, 2005). Accordingly, there exist a number of hermeneutic approaches to media text analysis with special focus on ‘emergent meaning’ as an aesthetic quality. Yet, the problem of investigating emergent phenomena from a perception, reception or cognitive perspective remains to be solved. Admittedly, there exists an experimental research tradition in the cognitive sciences of music that agrees with Chion’s thesis of *added value* (1994), and takes the pioneering work of Marshall & Cohen (1988) as a starting point (e.g. Hung, 2001). However, ‘added value’ is not consistent with ‘emergent meaning’; indeed, true emergence is *not additive* in nature, considering that a multimodal ‘whole’ cannot be deduced from the discrete component parts.

Now, if the concept of emergence is applied on the free-induction part of the study, the results might be a bit disappointing because generally the data do *not* point towards ‘emergent qualities’ in the audio-visual versions. Rather, the music and moving images seem to reinforce each other mutually, and therefore the audio-visual ‘wholes’ transcend the parts. As a consequence I might choose other kinds of audio-visual stimuli (containing more signifying elements) for a future experiment. Additionally, I might use physiological measuring data.

References

Aaker, J.F. (1997). ‘Dimensions of Brand Personality’, *Journal of Marketing Research, 34*, 347-356.

Burn, A., & Parker, D. (2003). *Analysing Media Texts*. London: Continuum.

Chion, M. (1994). *Audio-Vision. Sound on Screen*. New York: Columbia University Press.

Cook, N. (1998). *Analysing Musical Multimedia*. Oxford: Oxford University Press.

Ellis, R. J.; & Simons, R. F. (2005) ‘The Impact of Music on Subjective and Physiological Indices of Emotion While Viewing Film’. *Psychomusicology, 19,* 15-40.

Hull, G. A., & Nelson, M. E. (2005). ‘Locating the Semiotic Power of Multimodality’. *Written Communication*, 22(2), 224-261.

Hung, K. (2001). ‘Framing Meaning Perceptions with Music: The Case of Teaser Ads’. *Journal of Advertising,* 30(3), 39-49.

Libscomb, S. D., & Kendall, R. A. (1994) ‘Perceptual Judgement of the Relationship between Musical and Visual Components in Film’. *Psychomusicology, 13*, 60-98.

Marshall, S. K., & Cohen, A. J. (1988). ‘Effects of Musical Soundtracks on Attitudes toward Animated Geometric Figures’. *Music Perception*, 6(1), 95-112.

Metz, Ch. (1974). *Film Language: A Semiotics of the Cinema*. New York: Oxford University Press.

Middleton, R. (1990). *Studying Popular Music*. Milton Keynes: Open University Press.

Pauli, H. (1976) ‘Filmmusik: Ein historisch-kritischer Abriß’. In H.-C. Schmidt (Ed.), *Musik in den Massenmedien Rundfunk und Fernsehen: Perspektiven und Materialien*, pp. 91-119. Mainz: Schott.

Russel, J.A. (1980). ‘A Circumplex Model of Affect’. *Journal of Personality and Social Psychology, 39*(6), 1161-1178.

Tagg, Ph. (1989) ‘An Anthropology of Stereotypes in TV Music?’ *Svensk tidskrift för musikforskning*, *19*, 19-42.