

Investigating Multimodal Emergence through Experimental Audience Research

Paper for the research seminar and workshop on empirical methods for studying sound/music in audiovisual contexts

Anders Bonde

Introduction

The present study is motivated by a common notion among analysts of film music or film sound as well as theorists within the field of multimodal research; the notion that the perception of audio-visual media products, in which different semiotic resources or *modes* are combined, brings forth meanings that transcend the expressive and semantic potentials in the modes themselves.¹ As far as that is concerned, multimodal meaning is a phenomenon of *emergence*; i.e. an irreducible, coherent whole that cannot be grasped only by taking out the constituent modes for mutual comparison. Rather, the every single mode has to be analyzed and compared with the whole that occasionally seems to have a *downward causal effect* on the modes. Of course, this has often been a matter of qualitative hermeneutic analyses of art products with aesthetic value, but the problem is to investigate the phenomenon in a reception perspective, and that has not been carried out by anybody yet, as far as I am aware of. Admittedly, there exists an experimental research tradition in the field of music cognition that agrees with Michel Chion's thesis of "added value" (Chion 1994, 5-9) and takes the pioneering work of Marshall & Cohen (1988) as a starting point. However, the concept of added value is not consistent with emergence; that is, true emergence is not additive in nature since the multimodal whole cannot be deduced from the expressive and semantic potentials of the discrete component modes themselves. Therefore, I believe the experimental research tradition can be developed further where ideas about wholes and parts as known from emergentism (as well as hermeneutics) are incorporated.

For the present study, when testing *moving pictures, sound* (or music) and *moving pictures plus sound/music* separately, I have as a starting point decided to use self-reports as methodological framework and source of evidence while leaving physiological experiments for future research. The study was prepared in mid-February until mid-April 2010 by picking out audio-visual stimuli for test material, working out questionnaires, defining subjects, deciding the test setup with its different conditions and, off course, making the necessary appointments with people involved helping to realize the study. When all prepared for the study, we (the present writer and two colleagues) made in April 15 a *pretest* with the purpose of examining the research design. After a few but significant changes the main study was about to be initiated. In the following I will describe the experimental research design in detail

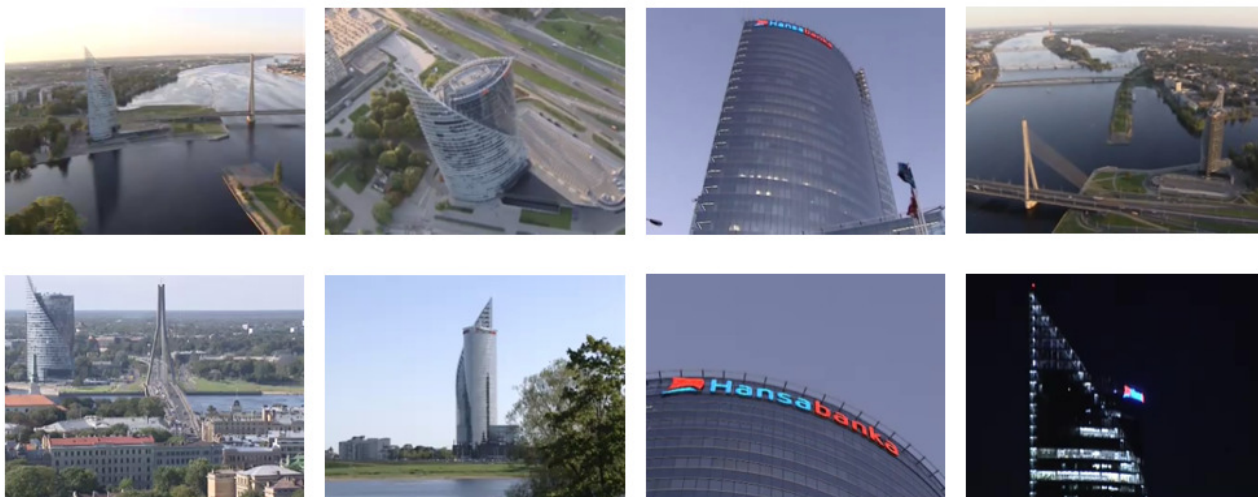
¹See for instance van Leeuwen (1985), Chion (1994), Cook (1998), Hung (2001), Burn & Parker (2003), Hull & Nelson (2005), Kress & van Leeuwen (2006), Langkjær (2006), Graakjær (2008) and Bonde (2009).

including the production of test material, types of stimuli and test conditions, choice of subjects and test setup as well as testing and reporting procedure.

Generating test material

The test material was a 31 sec. video featuring the Hansa Bank building in Riga, Latvia (cf. Illustration 1). The video, which was downloaded from the video-sharing website YouTube,² consists of a sequence of shots and zoom effects showing the building from different angles and perspectives, and a soundtrack of electronic music without lyrics. It is shot from air balloon by a Latvian guy and appears to be made as part of a documentary about Wooden Architecture of Riga. He offered the video for Hansa Bank like a promotional video, though the banking company has never used it.³

Illustration 1



Four undergraduate students in musicology at Aalborg University have manipulated the video, creating alternative musical tracks including sound effects. They were chosen by draw among seven students who showed interest in producing music and sound for the study. Each student was paid around DKR. 500. As working source they were given only the moving pictures (not the sound), and they were told to produce a soundtrack as if they were making music for a TV commercial. To ensure a variety of soundtracks, I arranged individual meetings with the student where we came to an agreement regarding musical style. Two of the students made also an alternative version even though they were not told to do so. In the end I chose 6 versions for the test, including the original version and one of the alternative versions.

The choice of test material has methodological and pragmatic reasons: To begin with, the video is rather simple due to a limited number of modes (only moving pictures and music; no texted messages, dialogue or voice over). Furthermore, the video portrays no human faces (in

²<http://www.youtube.com/watch?v=v9kC0qnDry0&feature=related>

³According to e-mail correspondance, August 6 (2010).

fact, there are no people at all!), and accordingly, no human emotions are communicated. This might be an advantage; with only two modes it is easier to know what is measured. Secondly, the film depicts a foreign (not Danish) banking company that is almost certainly unknown to the majority of the Danish population. That makes another advantage seeing that any eventual knowledge and viewpoints about the banking company do probably not need to be considered. Thirdly, the absence of dialogue and voice over simply makes it easy to change the soundtrack without losing important communicative elements.

The generated test material has one thing in common compared with the teaser films used in a study by Hung (2001) who has also used alternative musical tracks to the same pictures, and that is the absence of real sound and texted messages. Despite the fact that the “polysemous” (ibid., 40) montage flow with non-related shots of customers and employees in Raffles Shopping Center (Singapore) differs considerably from the Hansa Bank video with regard to its solely focus on the outward architecture of the building, I imagine that the videos share the feature that *musical style* becomes very important for the resultant meaning. This can perhaps, on the other hand, be a problem for the creation of multimodal emergence seeing that musical meaning is projected on the visual meaning and not vice versa. In Hung’s study, it is exactly the point to demonstrate such matter, but that is not the case in the present study, and therefore the test material can prove unsuitable for the purpose anyway. However, the answer to this remains to be made.

Stimuli and test conditions

Since we are interested in testing multimodal emergence through the interaction of visuals and music, it makes sense to distinguish between 3 different types of stimuli:

The audio-visual totality (1)

The visuals only (2)

The music and sound only (3)

For the pretest we used all three types derived from the original video saving all the alternative aural versions and audio-visual combinations for the study. Using such kind of differentiated stimuli is consistent with Marshall & Cohen (1988) and Hung (2001). What is important to notify, however, is that we will primarily focus on the comparison of (1) with (2) and (3) respectively, and secondarily the comparison of (2) with (3). Speaking of multimodal emergence, it makes sense first of all to compare every single modality with the multimodal whole because only in that way can we know if and how the perception of the audio-visual totality might be influenced differently by the visuals and the music/sound. For the same reason, just as it seems natural to ask the same questions regarding the *video* on the basis of (1) and (2), we have decided also to ask the same questions regarding the *music* on the basis of (1) and (3). In other words, there are *two* test conditions linked to stimulus (1); by presenting an audio-visual totality to a group of subjects, we will not only get information

about the video but also the music specifically. That makes a total number of 4 test-condition types, and due to the 6 versions, each with different music, the number of test conditions is 19 (see below).

- A: Stimulus (1) with a questionnaire regarding the film (6 test conditions)
- B: Stimulus (2) with a questionnaire regarding the film (1 test condition)
- C: Stimulus (3) with a questionnaire regarding the music (6 test conditions)
- D: Stimulus (1) with a questionnaire regarding the music (6 test conditions)

Subjects and test setup

12 graduate students in the field of communication participated in the pretest as part of an exercise after a lecture. They were divided in three groups with four students per group, each group assessing *only one stimulus*. The stimulus was presented three times, and after each presentation they were told to respond to it, at first freely and afterwards according to certain criteria (see below). For the study I have chosen the same procedure, also with a single stimulus per group of subjects. However, this time I have addressed high-school students at the Gymnasium and The Higher Preparatory Examination. The choice is pragmatically motivated: It has been rather attainable to visit high-school classes and carry out experiments like this, especially when I have offered in addition a lecture with some learning potential for the students. Hence, at the present moment I have visited 18 out of 19 classes in 6 different schools around The North Denmark Region, in which I have been allowed to spend a lesson (45-70 min.) with the students, running the test and then lecturing. Of course, I could have used a web-based questionnaire, which would have been much less time-consuming. However, I found it more valuable being able to guarantee a homogeneous group of subjects, all situated in the same context while participating in a controlled testing environment.

In that matter I consider the study as a kind of mix between a structured laboratory experiment and a structured field study. I have decided *not* to isolate the students from each other letting them do the test one by one in a room. Rather, I have presented the stimulus in the classroom (the visuals via a projector and the sound via loudspeakers) with the students sitting on their chairs. This is, on the other hand, not an attempt to simulate the average youth's everyday perception of TV commercials, nor is the film presented as a TV commercial (or the soundtrack as music from a TV commercial). Even though the film might *appear* as a TV commercial, it is taken out from its natural context (as a single spot among others), and the subjects are probably paying much more attention to it in a socially situated context of a teaching-learning milieu than if they were sitting in their homes watching TV. This, I believe, is not a shortcoming seeing that I intend to investigate what moving pictures and music *can* do together, and particularly, *if* and *how* they are able to interact in the perception of the listener and perhaps bring about the phenomenon of multimodal emergence. Hence, an attentive audience is needed. This is also why I have decided not to use more than one stimulus per class. The attention of an audience (and high-school students in particular!) is a

“vulnerable” matter; when presented with not only one but several stimuli after one another, the awareness of the subjects can be easily weakened.

Testing and reporting procedure: Free induction and questionnaires

A fundamental point in the testing procedure was *not* to inform the subjects in advance about the actual purpose of the study. Therefore, I have withheld all information about my hypotheses regarding multimodal emergence and the importance of musical style concerning meaning making in audio-visual media. The idea was to initiate the test by letting the subjects watch the video (or listen to the music) and after that write down what came into their mind (in 2 minutes) without being influenced by too much except by the stimulus itself.⁴ For the same reason, in all classes I have carried through the test straight away after a very few introductory remarks about who I was and what was going to happen in the next 20-25 minutes. The free-induction model as described above is quite similar to the procedure used by Philip Tagg (1989), and the reason for choosing such kind of self-report methodology, in general, is simply that “each answer has greater cultural and symbolic significance” when “created with the music [or video, *ed.*] as sole stimulus and not with the aid of ready made alternatives” (*ibid.*, 23).

The next part of the study is inspired by Hung (2001). Given the fact that I intend to compare the results from the 3 types of stimuli, it seems reasonable to use a standardized questionnaire, not considering the differences between the stimulus types, apart from the necessary adjustments regarding the wording, sometimes using “the video” and sometimes “the music”. Secondly, I have in the questionnaires focused on semantic-associative issues that seem to fit all three types of stimuli. Correspondingly, I have created two questionnaires, of which the first one (Questionnaire 1) contains 16 categories with regard to *emotion* while the second one (Questionnaire 2) addresses *brand image* and includes 27 categories. For every single category, concerning emotion or brand image, there are 4 possible answers: *No, on the contrary*; *No*; *Yes, to some extent*; and *Yes, very much*. The two questionnaires are illustrated in Appendix 1 and 2.

The 16 categories in Questionnaire 1 are inspired by Batra & Ray (1986), and so are the concept of emotion that is viewed as an inclusive label encompassing *affect*, *mood* and *feeling* (i.e. basic mental reactions). Additionally, the categories are adjusted according to Russel’s (1980) “circumplex model” that describes the relation between arousal and valence. That is to say, I have tried to pick out emotional categories representing all four combinations of arousal (high/low) and valence (positive/negative) roughly equal (cf. Table 1). Signifying *brand image*, though, is another matter being less associated with the subjects’ personal feelings. The shaping of the 27-item Question 2 is somewhat inspired by the five dimensions of brand

⁴Of course, there are lots of influential elements in the classroom: the unusual situation including my own presence, the student’s mutual communicative interactions, their present psychological state or mood etc.

personality (*Sincerity, Excitement, Competence, Sophistication* and *Ruggedness*) with its total of 15 facets as suggested by Jennifer Aaker (1997, 351), but generally, the chosen categories are primarily adjusted to fit the test material. As a result, the fifth dimension (*Ruggedness*) is not present while, on the other hand, images with negative emotional content are added to make a balance. However, as for brand image, the distinction between positive and negative valence might not be as clear as with regard to emotion (cf. Table 2).

Table 1

| <i>The video/music made me feel ...</i> | | | |
|---|-------------------------|--------------------|-------------------------|
| <i>High arousal</i> | stimulated | | annoyed |
| | enthusiastic | | repelled |
| | engaged | | confused |
| | moved | surprised | worried |
| <i>Low arousal</i> | in a good mood | wondered | uncertain |
| | satisfied | | sad |
| | relaxed | | bored |
| | | | |
| | <i>Positive valence</i> | <i>Neither nor</i> | <i>Negative valence</i> |

Table 2

| <i>The video/music makes an impression of...</i> | |
|--|-------------------------|
| sincerity | dishonesty |
| down to earth | irresponsibility |
| tradition | deceitfulness |
| simplicity | unreliability |
| accountability | naivety |
| stability | failure |
| trustworthiness | complacency |
| intelligence | unoriginality |
| success | poverty |
| self-realization | normalcy |
| inventiveness | anonymity |
| trendiness | |
| independence | |
| classiness | |
| glamour | |
| humor | |
| <i>Positive valence</i> | <i>Negative valence</i> |

Preliminary discussion of research design

As mentioned earlier, the 19 test conditions are almost completed with only one left. At the present I have not had time yet to analyze the data and make comparisons of results between test conditions, though I plan beginning the work during October and November 2010. Here, I assume that the most profitable approach will be the testing of hypotheses for verification (or refutation) concerning assertions about the influence of the constituent modes on the audio-visual totality and the other way round, instead of trying to extract patterns from the data. Nevertheless, I have made some preliminary analysis of the results from one class (labeled "1.y") exposed to the video containing alternative music (test condition A), and that will be a departure for a final discussion and reflection on some methodological problems about the questionnaires (involving both content and structure).

Concerning the 1.y results from Questionnaire 1, it appears that the answers tend to be negative while the results from Questionnaire 2 reflect a wider spectrum from the denying to the confirming (cf. Table 3-4). From this, it might be concluded that the categories of emotion are less consistent with the subject's feelings than seems to be the case as regards the categories of brand image. There might be several reasons to this, but I guess that the subjects are likely to be emotional unaffected by the video because of the absence of human faces. I imagine it might be psychologically easier to evaluate what impression a film (or music) makes than to put a word on personal feelings that are consistent with a list of predefined categories. As noted by Matthias Bode (2007, 13), "there is a difference between identification of feelings and actually *feel* those feelings." Perhaps, on the other hand, the emotional categories suit better for describing music. That remains to be concluded.

Table 3

| | <i>Category</i> | <i>Observed minimum</i> | <i>Observed maximum</i> | <i>Average</i> | <i>Number of subjects</i> |
|-----------------------------|-----------------|-------------------------|-------------------------|----------------|---------------------------|
| The video made me feel .../ | enthusiatic | 1,00 | 3,00 | 2,00 | 25 |
| The music made me feel ... | moved | 1,00 | 3,00 | 2,00 | 24 ^(*) |
| | repelled | 1,00 | 3,00 | 2,00 | 25 |
| | sad | 1,00 | 3,00 | 2,04 | 25 |
| | surprised | 1,00 | 3,00 | 2,04 | 25 |
| | uncertain | 1,00 | 4,00 | 2,12 | 25 |
| | stimulated | 1,00 | 3,00 | 2,12 | 25 |
| | engaged | 1,00 | 3,00 | 2,16 | 25 |
| | annoyed | 1,00 | 4,00 | 2,20 | 25 |
| | worried | 1,00 | 4,00 | 2,21 | 24 ^(*) |
| | satisfied | 1,00 | 3,00 | 2,24 | 25 |
| | in a good mood | 1,00 | 4,00 | 2,24 | 25 |
| | bores | 1,00 | 4,00 | 2,48 | 25 |
| | relaxed | 1,00 | 4,00 | 2,52 | 25 |
| | confused | 1,00 | 4,00 | 2,56 | 25 |
| | wondered | 2,00 | 3,00 | 2,60 | 25 |

Table 4

| | <i>Category</i> | <i>Observed minimum</i> | <i>Observed maximum</i> | <i>Average</i> | <i>Number of subjects</i> |
|---------------------------------------|------------------|-------------------------|-------------------------|----------------|---------------------------|
| The video makes an impression of .../ | poverty | 1,00 | 2,00 | 1,36 | 25 |
| The music makes an impression of .../ | humor | 1,00 | 2,00 | 1,64 | 25 |
| | failure | 1,00 | 3,00 | 1,88 | 25 |
| | tradition | 1,00 | 3,00 | 1,92 | 24 ^(*) |
| | down to earth | 1,00 | 3,00 | 1,96 | 25 |
| | unoriginality | 1,00 | 4,00 | 2,00 | 25 |
| | irresponsibility | 1,00 | 3,00 | 2,08 | 25 |
| | unreliability | 2,00 | 4,00 | 2,24 | 25 |
| | dishonesty | 1,00 | 4,00 | 2,28 | 25 |
| | anonymity | 1,00 | 4,00 | 2,32 | 25 |
| | deceitfulness | 1,00 | 4,00 | 2,36 | 25 |
| | simplicity | 1,00 | 4,00 | 2,40 | 25 |
| | inventiveness | 1,00 | 4,00 | 2,42 | 24 ^(*) |
| | normalcy | 1,00 | 4,00 | 2,48 | 25 |
| | naivety | 1,00 | 4,00 | 2,60 | 25 |
| | self-realization | 2,00 | 4,00 | 2,62 | 24 ^(*) |
| | sincerity | 1,00 | 4,00 | 2,68 | 25 |
| | trustworthiness | 2,00 | 4,00 | 2,72 | 25 |
| | complacency | 2,00 | 4,00 | 2,76 | 25 |
| | glamour | 1,00 | 4,00 | 2,80 | 25 |
| | stability | 2,00 | 4,00 | 2,80 | 25 |
| | accountability | 2,00 | 4,00 | 2,84 | 25 |
| | independence | 1,00 | 4,00 | 2,88 | 25 |
| | trendiness | 1,00 | 4,00 | 3,08 | 25 |
| | classiness | 2,00 | 4,00 | 3,08 | 25 |
| | intelligence | 2,00 | 4,00 | 3,20 | 25 |
| | success | 2,00 | 4,00 | 3,44 | 25 |

Another issue is the structure of the multiple-choice format that appears to have a couple of weaknesses. Firstly, a null answer is missing. With no answers reported, I cannot be sure if the subject has simply overlooked the category or if he/she has not felt able to make a decision. Secondly, the four possible answers are not represented in separate boxes but written next to each other in a table looking like a continuous answer scale. This is not appropriate because some of the subjects have put a mark in between two answer possibilities, and therefore being invalid.

There are probably other issues and problems worthy of discussion, not only regarding the questionnaires but the research design in general, and I will look forward to do so at the forthcoming research seminar in October 13-14.

References

- Aaker, J.F. (1997). Dimensions of Brand Personality”, *Journal of Marketing Research*, 34, 347-356.
- Batra, R., & Ray, M.L. (1986). Affective Responses Mediating Acceptance of Advertising, *Journal of Consumer Research*, 13, 234-249.
- Bode, M. (2007). The Sounds of Selling, *BrandBase*, 21, 12-14.
- Bonde, A. (2009). On the Commercialization of Shostakovich’s ‘Waltz No. 2’: A Case Study of Textual, Contextual and Intertextual Meaning of Music. In N. J. Graakjær & Ch. Jantzen (Eds.), *Music in Advertising: Commercial Sounds in Media Communication and Other Settings* (pp. 141-167). Aalborg: Aalborg University Press.
- 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.
- Graakjær, N. J. (2008). *Musik i tv-reklamer. En tekstanalytisk undersøgelse*. Aalborg University.
- 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.
- Kress, G., & van Leeuwen, Th. (2001). *Multimodal Discourse. The Modes and Media of Contemporary Communication*. London: Arnold.
- Kress, G., & van Leeuwen, Th. (2006). *Reading Images: the Grammar of Visual Design*. London: Routledge.
- Langkjær, B. (2006). Mediernes lyd – En multimodal analysemetode. *Mediekultur*, 40, 14-26.
- Marshall, S. K., & Cohen, A. J. (1988). Effects of Musical Soundtracks on Attitudes toward Animated Geometric Figures. *Music Perception*, 6 (1), 95-112.
- Russel, J.A. (1980). A Circumplex Model of Affect, *Journal of Personality and Social Psychology*, 39(6), 1161-1178.
- Tagg, Ph. (1989) An Anthropolgy of Stereotypes in TV Music? *Svensk tidskrift för musikforskning*, pp. 19-42.
- Van Leeuwen, T. (1985) Rhythmic Structure of the Film Text. In T. A. van Dijk (ed.) *Discourse and Communication: New Approaches to the Analysis of Mass Media Discourse and Communication*. Berlin: de Gruyter.