akademisk kvarter
academic quarter

Volume 16. Autumn 2017 • on the web

Studying the Aesthetics of Images and Advertising Films

Combining Systemic-Functional Grammar and Audience Physiology

Anders Bonde *is associate professor at the Department of Communication and Psychology, Aalborg University, and member of the knowledge group MÆRKK. Much of his research is dedicated to experimental aesthetics including the use of affecting-computing methods in empirical media and art studies, with a particular focus on music and sound.*

Birger Larsen *is professor in Information Analysis and Information Retrieval at the Department of Communication and Psychology, Aalborg University Copenhagen. He is head of the AAU knowledge group on Communication and Information Studies and leads the research group on Information Analytics.*

Abstract

The notion, that aesthetic objects are not only things but also *does* something to those who engage with them, is consistent among various researchers in the arts and humanities; and also that this is not primarily a matter of interpretative symbolic meaning in the fields of semiotics and hermeneutics, but may rather be associated with sense perception, such as arousal, vigilance, fascination and appraisal (Marković 2012). In this position paper, we sketch out a conceptual approach for the study of art by combining, on the one hand, the systemic-functional and social-semiotic approach to textual analysis of grammar (Kress & van Leeuwen 2006, van Leeuwen 1999), and on the other hand, attentional and affective-computing techniques for measuring audience response.

Keywords aesthetics, sense perception, systemic-functional grammar, eye tracking, electrodermal activity.

Introduction

It is often stated that an aesthetic experience of art will always be a personal matter; that is, something that cannot be measured or defined by general rules. In the same vein, the aesthetic experience of art has often been subjected to hermeneutics, phenomenology and cultural studies, giving preference to qualitative non-directive interviewing and observation methods. Concurrently, empirical and quantitative approaches have been heavily criticized as reductionist and insufficient for understanding the nature of an aesthetic experience, relying on a number of misconceptions about art (see, e.g., Wheelwell 2000). It seems, though, that the problem lies in the vague or even absent definition of the concept of 'aesthetic experience', at least in the field of experimental aesthetics where some clarifications seem appropriate. For instance, there are at least three overall characteristics or aspects to consider in that matter; that is, the *attentional* aspect, being a state of high arousal and vigilance in which the spectator or listener is focused on and fascinated with an aesthetic object; the *cognitive* aspect, being a state of high appraisal of the semantic, symbolic and imaginative referents of an aesthetic object; and the *affective* aspect, being a strong feeling of unity with the object of aesthetic fascination and appraisal (Marković 2012, 3).

In this paper, we will concentrate on the attentional and affective aspects in relation to aesthetic content. More specifically, we sketch out a conceptual approach for investigating correspondences between, on the one hand, social-semiotic and systemic-functional grammar analysis of visual and audio-visual aesthetic objects (Kress & van Leeuwen 2006, van Leeuwen 1999), and on the other hand, eye-tracking (ET) and electrodermal-activity (EDA) studies of subjects exposed to the objects, the latter being examples of popular, simple and relatively inexpensive tools. Introduced in detail below, the aesthetic objects include a pair of food photographs and a Japanese snow-tyre TV commercial.

The reason for combining text analysis and experimental audience-testing research seems evident: One thing is to complement attentional and affective computing with cognitive data (e.g., introspective or retrospective think-aloud protocols), thus obtaining missing insights into people's minds while or after looking at an aesthetic object; another thing (being the core matter in the present paper) is to understand potential connections between the atten-

tional and affective responses and the content of the aesthetic objects that elicit them. The latter, straightforward as it might seem, happens to be, by and large, a rarity in empirical-experimental research in which aesthetic objects (used as stimuli) are seldom analysed systematically for structure, function and meaning. Of course, we cannot redress such insufficiencies in a short paper like the present one. Instead, our aim is to state a position by providing analytical examples of what might be achieved through a limited but significant choice of aesthetico-semiotic and physiological methods used in combination.

Grammar analysis and experimental aesthetics

Why do we find social-semiotics and the systemic-functional approach useful? For one thing, bringing aesthetics under semiotic theories is a familiar and well-proven strategy in terms of analysing art or artistic endeavour (cf. Rudner 1951; Morris 1939), and there have been previous attempts to do so in the field of social semiotics (cf. O'Toole 1990; Kress 2010). Moreover, the approach, which is originally based on theories of Halliday (1978, 1985), offers a high level of detailing, formalization and abstraction that appear fit for method-parallelization processes (Junk 2011) with quantitative data from ET and EDA monitoring. Finally, the strong emphasis on the significance of *context* for determining how "images relate the people, places and things they portray to each other, so as to form coherent representations", and "[w]hat kind of 'interpersonal resources' do the images use to create a relation between the image and the viewer" (Van Leeuwen 1999, 190), seems advisable when analysing sense or meaning making.

In order to provide an analytic-descriptive system for characterizing any visual image on the basis of context-dependent semiotic properties, Kress and van Leeuwen have defined nine 'modality markers' (2006, 160–162, 264); that is, mutually independent parameters, each being characterized as continua with an infinite number of articulatory levels between two opposite extremes, from *maximum simplicity* (i.e., a high level of abstraction) to *maximum complexity* (i.e., a high level of detail), and within amplification or reduction might take place (cf. Figure 1)[1]. Analogously, van Leeuwen (1999) has identified eight markers in the domain of sound (cf. Figure 2). The basic idea in both cases is that modality, as defined as the

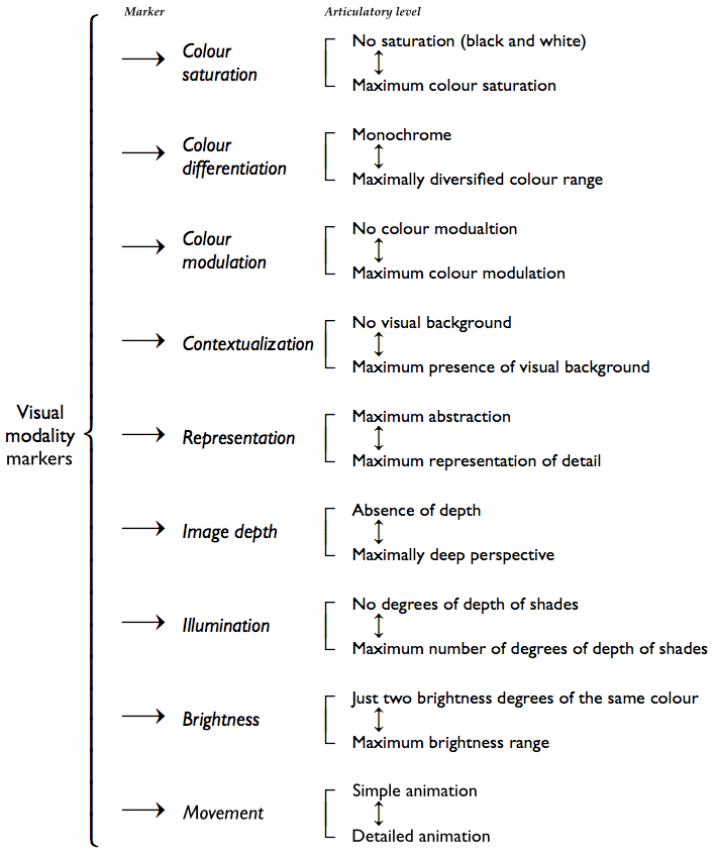

Figure 1. Visual modality markers (Kress & van Leeuwen 2006).

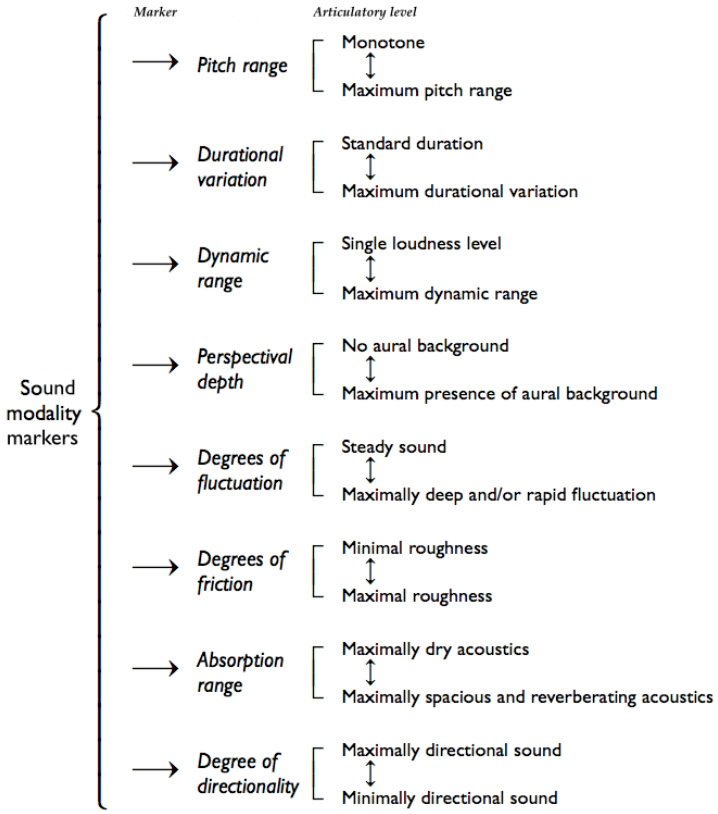

Figure 2. Sound modality markers (van Leeuwen 1999).

amount of reliability or credibility assigned to any given image or sound, "is realized by a complex interplay" of settings according to a number of individual markers (Kress & van Leeuwen 2006, 163); and by the systemic-functional or structural approach, the authors offer an all-inclusive, transparent and logical framework for mapping out socially constructed resources for the communication of meaning through images. Moreover, and what is not explicated directly, the generic attributes of the modality markers may be considered as carriers of *aesthetic* meaning having influence on attentional information processing and affective arousal; and this is why the markers are particularly useful in inter-paradigmatic research combining arts (or humanities) and science as suggested in the present paper. The markers simply form a set of explanatory (inde-

pendent) content variables, which can easily be aligned with (quantitative) data from experimental audience research.

As for the *contextual matter*, which is crucial in the field of social semiotics, Kress & van Leeuwen (2006) distinguish between technological, sensory, abstract and naturalistic kinds of 'coding orientation', the latter being a concept that is originally developed by Bernstein (1971, 1981) with regard to socio-cultural and situational context, that is, "sets of abstract principles which inform the way in which texts are coded by specific social groups, or within specific institutional contexts" (Kress & van Leeuwen 2006, 165). That means that the modality of an image (or a sound) may differ from one context of perception to another, either across socio-cultural boundaries, or within the same socio-culture but being conditioned on the actual situation. For instance, the modality of a densely-coloured painting in an art museum will most likely be estimated higher than the modality of an equally densely-coloured an architect drawing (even by the same perceiver).[2]

To enable the combination of analytical findings with ET and EDA data, we use distinct scale values rather than continua. Besides, due to the position-stating aim of the present paper, we have, for the sake of simplicity, chosen five-point scales, though a finer-grained scaling would probably be preferable in a more exhaustive analysis. Likewise, for simplicity reasons, while recognizing the plurality of visuals in moving images, we have in the analysis of the commercial applied the visual modality markers in relation to the total sequence of images. This makes sense to the extent that the configuration of modality settings is reasonably invariable during the motion. Similarly, as for the auditory domain, we have applied the sound-modality markers in relation to the sound part in its totality rather than, as implied and exemplified by van Leeuwen (1999, 184–185), in relation to distinct sounds. Accordingly, in the following analytical sketch, the configuration of settings within modality markers nos. 1–8 (cf. Figure 1) will be pivotal points when subsequently aligning the findings with ET data. After that, we exemplify briefly how a similar text-analytical approach might be extended to the analysis of moving images (incorporating all nine modality markers in Figure 1) and sound (incorporating the eight modality markers in Figure 2), and how the results might be aligned with EDA data.

Analysis of the food images

The photographic pair was selected on the basis of a search for images with a creative commons license on the popular photo-sharing site Flickr. The idea behind the photo selection originates from an experimental ET study, which was designed and undertaken by a group of students on an elective course in *Applied Aesthetics* as part of the Master's degree in Communication at Aalborg University. The students tested a number of photo pairs displaying a vegetarian and non-vegetarian version of the same type of food. The aim of the study was to identify whether viewers tend to focus on the one or the other kind and, additionally, if there were any attention differences between male and female viewers. In the present paper we have, in similar fashion, picked out two photos of a grilled burger and a vegetarian (or 'veggie') burger, respectively (cf. Figures 3 and 4); and as a supplementary set of descriptors, we include content information in form of the eight modality markers (cf. Figure 1).

Figure 3. Grilled (meat) burger.
Photo courtesy of Alpha (https://flic.kr/p/4gRyYm)

Figure 4. Veggie burger.
Photo courtesy of lara604 (https://flic.kr/p/67dPFp)

When studying the two burger photos, one might come across a number of compositional similarities. To begin with, each photo reveals a kind of gourmet-type burger characterized by a home-baked bun, green salad, cheese and slices of tomatoes and red onions, as well as a grilled beef or, in the case of the veggie burger, a

black-bean paste, all held together with a strong toothpick. Secondly, the burgers are both delicately arranged on shiny white plates and shown in an extreme close-up shot on a blurred background. There are, however, also some differences to notice. Aside from the alternative choice of 'protein component' (black-bean paste instead of grilled beef), the veggie burger has been split in two halves with the right piece placed slantwise on the left piece, while the grilled burger remains in one piece. Furthermore, the veggie burger contains a layer of light-green sprouts between the salad and the upper part of the bun, and the colour of the (cheddar) cheese is dark yellow and not white as in the grilled burger. Lastly, while both being seen in central perspective (cf. *image depth*), the photographic angle of the grilled-burger photo is somewhat higher than the angle of the veggie-burger photo, where the front of the burger appears at eye level. This inevitably leads to a slightly higher level of *contextualization* in the grilled-burger photo, testified by the fact that one is able to identify a person's arm and the contours of (possibly) a class-

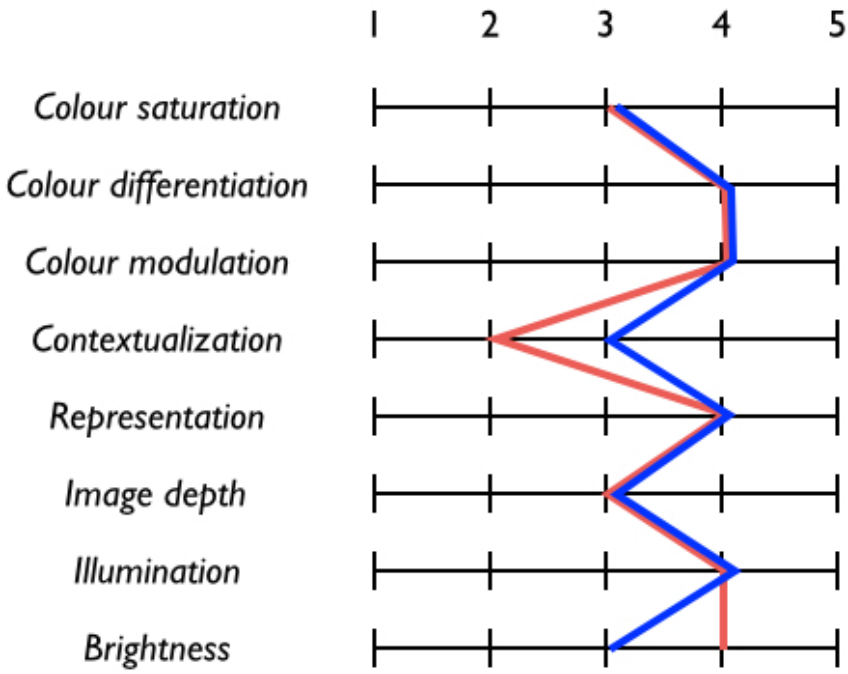

Figure 5. Modality configurations of the grilled burger photo (blue) and the veggie burger photo (red).

room floor on which furniture is set. As regards other modality markers and their respective settings, it might be worth concentrating on *brightness*, seeing that the difference between the lightest and darkest areas seem to be more profound in the veggie-burger photo, with a greater amount of deep blacks in the background and bright white flickers in the colours of the vegetables in the slanted piece to the right. In sum, according to the eight modality markers, the scale settings of the two photos might be illustrated as displayed in Figure 5, through which the (minor) modality differences can be observed and interpreted. For instance, the veggie-burger photo might be characterized as a little more 'artistic' and less 'natural' than the grilled-burger photo; compared to the grilled burger, the veggie burger tends to appear (but not entirely), what Kress and van Leeuwen describe as "more than real" or "hyper-real" (2006, 163–164).

The ET study

The use of eye tracking goes back to before 1900 (see, e.g., Huey 1898) and is an increasingly popular method of studying the attention of subjects on different stimuli, ranging from print and on-screen text, works of art, web sites, and recently also spatial analysis in 3D settings as, for instance, in museums (see Bergstrom & Schall 2014 and Bojko 2013 for an overview and Magnussen et al. 2016 for an example of a museum study). The use of eye tracking relies on the notion that human physiological capability to obtain visual evidence from the surroundings is limited by the structure of the human eye: high-acuity visual data can only be perceived from a very narrow visual angle at any given point (Land 2014). In addition, perceptual processing capacity is also limited, and thus there is a high correlation between the direction of a person's gaze and the focus of attention. Therefore, continuously tracing where a person is looking can offer rich information about what data is being sampled and used in visually guiding the activities people are involved in (Lauwereyns 2012). In the present study we used an low-cost eye tracker developed by The Eye Tribe[3], in the form of a bar placed below and in front of a 22 inch computer flat panel screen, which is a quite unobtrusive setting not too different from normal use of a desktop computer. The tracker samples both eyes of a subject at 60Hz. The open source Ogama software[4] was used to present stimuli and record and analyse gaze data. We created a single stimuli

slide containing the photos in Figures 3 and 4, and without instructions exposed this slide to subjects for 15 seconds while eye-tracking them. This short exposure was chosen to capture the immediate points of focus of the subjects.

In the following, we report on the results from two subjects, a male and a female, participating in this ET experiment. The purpose is to illustrate the type of data that can be collected, as well as to exemplify the kind of results and interpretations that can be based on this. Eye movements consist of so-called *fixations*, which are short periods of time where the eye rests on a single point of attention and can perceive visual data (typically more than 200–300 ms), and so-called *saccades*, which are quick jumps between fixations where visual input is reduced. In the 15 second exposure the subjects had 6–7 fixations. The eye-tracking software records both the duration and location of the fixations and these usually form the basis of the analysis and interpretation. Several different types of analysis can be performed.

Figure 6 shows the stimuli slide with an overlay of the *scan path* of the two subjects (the female in yellow and the male in green/dashed). The numbered scan path shows the order of the subjects' fixations, with lines drawn between them to indicate saccades, and the circle sizes represent fixation duration. We see that the female subject (yellow) has fairly many but short fixations. She first gazes at the veggie burger, then at the meat burger and then some longer at the veggie burger. At the meat burger, the prominent lower part the burger itself as well as the plate and person in the background was noticed. On the veggie burger, the background and in particular the bean paste was noticed. The male subject (green/dashed) has fewer, but much longer fixations (see also Table 1 below). He first focuses on the meat-burger filling and for a long time on the piece of bacon, and then moves on to the veggie burger focussing on the centre of the filling of both pieces of the burger. It is also interesting to note that the split veggie burger seems to attract attention to both halves in both subjects.

While scan-path visualisations are good for detailed studies of gaze behaviour, including flow and ordering and individual focus points of individual subjects, they become impractical with long exposures to stimuli or with many subjects. Here heat maps, such as the one illustrated in Figure 7, are more useful. Figure 7 shows a

Figure 6. Scan path visualisation (female = yellow, male = green/dashed).

Figure 7. Heat map aggregation of male subject fixations.

heat map where the fixation information is aggregated using a colour scale, where red and yellow indicate frequent/longer fixations and blue and purple colours indicate the least fixation activity. In the present example only one subject is included (the male), but the technique can be used on large numbers of subjects as a way of summarising a large amount of gaze data. Again we see the clear focus on the meat burger and in particular the bacon, as well as the dual focus on both halves of the veggie burger's filling. Taken altogether, the result might, for one thing, suggest a hypothesis that contextualization differences between the two burger photos (cf. Figure 5) lead to attentional differences among females; and, secondly, that a male audience might be expected to focus on meat rather than vegetables. Besides, the more decorative presentation of the veggie burger has attentional implications for both sexes.

Finally, we present an *area-of-impact* analysis, in which certain areas of particular interest are defined and various fixation statistics are calculated for these specific areas. In the present study we simply split the stimulus vertically screen in two, with a burger in each. Various areas can be defined including polygons and even overlapping ones. For a more thorough study, the selection of these areas could be informed by the analysis of the food photos above. Table 1 shows that the male subject has his first fixation on the meat burger and looked longer at it than the veggie burger both in total and in average per fixation. The female subject has 4 out of 7 fixations on the veggie burger, including the first one. She also looks longer on average and in total on the veggie burger. The absolute numbers are

Measure	Female	Male
First fixation (total number of fixations)	veggie (7)	meat (6)
Number of fixations (meat/veggie)	3 / 4	3 / 3
Average fixation duration (meat/veggie)	189 ms / 224 ms	933 ms / 522 ms
Total fixation duration (meat/veggie)	566ms (39%) / 897ms (61%)	2799 ms (64%) / 1565 ms (36%)

Table 1. Area of interest analysis.

of little interest (e.g., that the male has longer fixations overall) - it is the relative statistics that are of greater interest (e.g., that the female and males looked approximately ⅔ of the time on the veggie and meat burger respectively).

With only one subject in each gender group and a very short exposure to the stimuli, the presented eye-tracking results are only for illustration. However, they do demonstrate the potential for an interesting study of a larger group of subjects – as discussed below.

Analysis of the TV commercial

Released in December 2013, a Japanese TV commercial, promoting car tyres from the manufacturer Autoway (ostensibly suitable for snowy weather),[5] generated considerable media attention as 'the scariest ad ever' (see, e.g., Edwards 2013; Green 2013; Withnall 2013). Indeed, as warned by red text colour in the beginning of the commercial, and subsequently demonstrated by several unexpected scary moments by way of a sudden close-up shot of a ghoulish figure (complemented by dramatic sound effects), it is not for the faint-hearted. For that reason, due to its arousal-invoking content and hence its capabilities of (potentially) causing sweat-glands variations in the skin that are measurable, the tyre commercial was included as stimulus in an EDA experiment with the participation of 79 students from the Bachelor degree in Communication and Digital Media at Aalborg University. The experiment was conducted for didactic purposes and figured as a part of the curriculum in a prescribed fourth-semester course in *Aesthetics and Effect*. In the following, we will describe the content and story of the 'ad thriller', and after that we give an account of the design and results of the EDA study.

The story unfolds through the eyes of a driver and a passenger (both men) who are driving along a dark and snowy road at night in the middle of nowhere while having a low-voiced conversation. The viewing conditions are rather bad due to numerous snowflakes that are drifting by the windshield. At a certain point, the men catch the sight of a ghoulish-like figure (perhaps a young woman) in the middle of the road, wearing (it seems) only a thin nightdress. As the driver stops the car, keeping in proper distance from the figure (and probably waiting for her to step aside the road), she suddenly hits the windshield with a loud bang and appears with her eyes wide

open in front of the driver, and one immediately notices the ghoulishness and scariness of the woman due to her black-edged eyes and scarred left cheek. The windshield collision affects a ghoulish shout, and the shocked motorists start yelling and screaming while hazardously reversing away, probably without noticing that the woman is holding an open laptop with a texted message (in Japanese) saying: 'Have you put your winter tyres on?' However, the instinctive action of the driver clearly demonstrates the performance of the tyres; and as a final zoom-in clip of the laptop, the advertising message is revealed for a longer period of time to the sound of driving. The commercial ends with the Autoway logo and the company's web address, initiated by another bang sound.

Given that the spectator experiences the scary incident as if (s)he is the driver or is sitting in the car, thus perceiving the course of action from a central perspective (cf. *image depth*), the narrative possesses realistic and immersive qualities.[6] The shaky footage due to the hand-held ('subjective') camera, so typical for the aesthetics of horror films (see, e.g., *The Blair Witch Project*, 1999), contributes significantly to this accomplishment. Despite the artificially amplified "tension in the play between concealment and revelation" (Monnet 2015, 146), the filmic representation of *movement* appears quite realistic. Also, except for the ghoulish woman on the road, the 'naturalness' or 'realism' of the spooky night scenery with the absence of *colour saturation* and *colour differentiation* seems quite convincing. When driving in a car at night looking out the windshield (and not being outside in the dark), one is hardly able to distinguish any colours; only different shades of grey (cf. *colour modulation*) are visible, but even then the articulatory diversity is limited, as in regard of *illumination* and *brightness*, which is due to the contrast between the visible spot on the road that is illuminated by the car lights and the pitch-black surroundings. All such matters suggest a high modality in the domain of naturalistic coding orientation (Kress & van Leeuwen 2006, 165). However, otherwise unnatural appears the reduced level of *representation*. The image of the ghoulish woman abstracts from 'photographic naturalism' (2006, 161), since the figure on the road is never exposed in detail, nor in the zoom-in clip at the end.

Considering the auditory dimension, the sense of realism is supported according to a number of markers such as, for instance, *de-*

grees of friction, *degree of directionality* and *perspectival depth*. As for the former one, the grinding noisiness of the car's engine and the friction of the tyres on the road are fairly detailed, consistent with real-world experiences, while as for the latter two, the realism is established by a comparably low level of complexity. More specifically, the sounds are easily localized (cf. *degree of directionality*), that is, "pinpointed to a specific source" (van Leeuwen 1999, 177), and the 'flat' internal soundscape of the car's cabin is characterized by a low differentiation between foreground and background (cf. *perspectival depth*) due to a naturally occurring interference between to two men's conversation on the one hand and the exterior car sounds on the other. At the collision point, *pitch range*, *dynamic range* and, to a lesser extent, *durational variation*, are widened because of the bang on the windshield, followed by the men's emotional outbursts and the car reverse sounds (both the engine and interior beeps), though without affecting significantly the degree of friction and perspectival depth. All in all, similarly to the modality configurations of the burger photos (cf. Figure 5), the scale settings of the visual and auditory dimensions of the commercial might be illustrated as displayed in Figure 8.

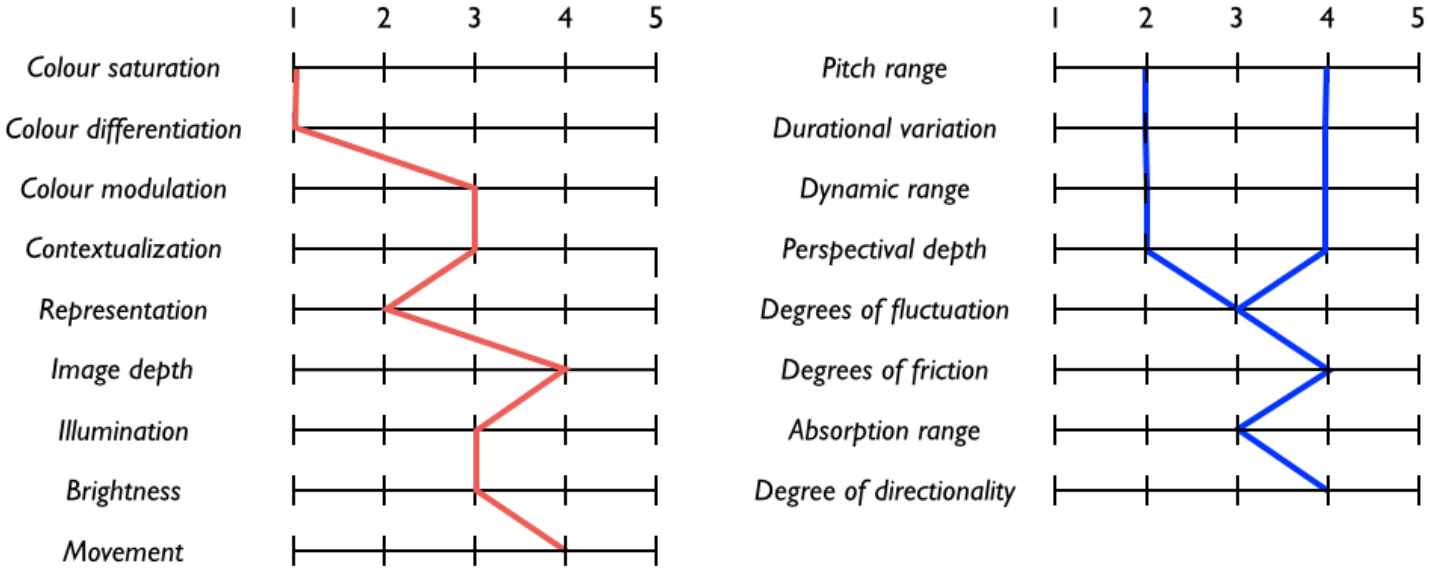

Figure 8. Visual and sound modality configurations of the commercial (red and blue, respectively).

The EDA study

There seems to be broad agreement among scholars in the fields of marketing psychology and consumer behaviour that affective-computing technologies for monitoring physiological signals, such as EDA, are highly relevant for studying audiences' experience of TV and web advertising. By providing an exact quantitative description of the electrical conductivity of the skin's moisture level (i.e., the state of sweat glands activated by the sympathetic nervous system), they may offer insight into emotionally induced reactions that are not immediately accessible to the researcher and perhaps neither "available to viewers' conscious awareness" (Ravaja 2004, 195). Additionally, the tools of EDA monitoring have become relatively simple and inexpensive to manage (Boucsein 2012, 1). However, being an indicator of the intensity of an emotional experience (or the state of physiological and psychological 'awareness'), EDA measures provide results only for one of two dimensions of affective experience (cf. Russell 1980), that is *arousal*, while revealing no information about *valence*, that is, the quality of the experience. Moreover, EDA measures are hardly of any use without being processed in combination with additional and explanatory information, as for instance (subjective) participation feedback or, in this case, stimulus annotations in the form of textual (visual and auditory) grammar configurations.

In the following, we report from the procedures and the results from an EDA experiment in which the Japanese tyre commercial was used as stimulus. 79 students watched the commercial while they were monitored by a wireless Sense Wear armband from Body Media attached by a Velcro strap on the backside of the upper left arm. The experiment took place in a dark media lab at Aalborg University, where the students were scheduled to participate in groups of three. The tyre commercial was incorporated in a 10-minute compilation including seven advertising or campaign films, which were specifically chosen for the experiment, given their emotional content and (presumed) capability of invoking affective response. In the following we shall, however, concentrate exclusively on the data derived from watching the tyre commercial. In this connection it should be noted that we chose for the experiment a modified and extended version, in which the zoom-in clip of the laptop and the subsequent bang sound is cross-faded with

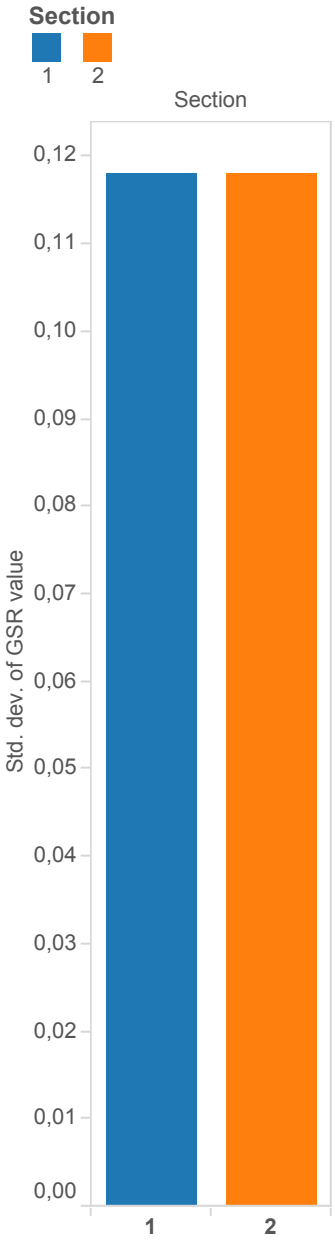

Figure 9: Between-subject variability in Sections 1 and 2.

a repetition of the scary sequence (car stopping followed by the collision and the reversing), which is once again cross-faded with a slow-motion replay of the same sequence (except for the reversing) before the final logo and bang sound. The choice was motivated by a wish to retain the intensity a bit longer so as to improve the ability to measure the differences of affective reactivity before and after the first shock effect. Accordingly, in the following paragraph we will account for the resulting data belonging to the two sections on each side of the (first) collision, that is Section 1 (0:00–0:18) and Section 2 (0:19–0:54).

When we compare EDA variability in Section 1 with EDA variability in Section 2, there are virtually no differences at all (cf. Figure 9). That simply means, though, that the general between-subject differences are rather constant regardless of the perceived media content and the evoked impressions. Thus, the two columns in Figure 9 describe the average EDA variability across all 79 students. When we, however, conduct an equivalent examination of the two sections on a within-subject level, it becomes very clear that Section 2 exceeds considerably Section 1 in far most cases, with only five exceptions, that is, in case of participant no. 8, 68, 84, 89 and 93 (cf. Figure 10). This result – which is further illustrated in a box-and-whisker plot, showing both the typical and atypical values of variability (cf. Figure 11) – might be interpreted as a clear indication that the level of affective arousal among the participants is markedly greater after the collision episode; and it confirms what one would expect in the light of the scary close-up confrontation and shout of the ghoulish figure followed by the high-intensity exposure of the emotional reactions of the motorists. Indeed, the significance of the shock effect should not be underestimated. Yet, it seems also reasonable to assume that the altered configuration of sound-modality markers (cf. Figure 8), which is due to the intensified soundscape, plays a major role in the elicitation of arousal increase.

That said, one should bear in mind that Section 2, with its repetitive content, hides the affective impact of the first shock effect and its influence on the replayed shock effects.[7] Thus, it raises the question whether the increased EDA variability in Section 2 may be attributable to within differences before the extension (0:19–0:36) or within the extension (0:37–0:54), or to differences between these two subsections. The latter is illustrated as '2a' and '2b' in Figure 12,

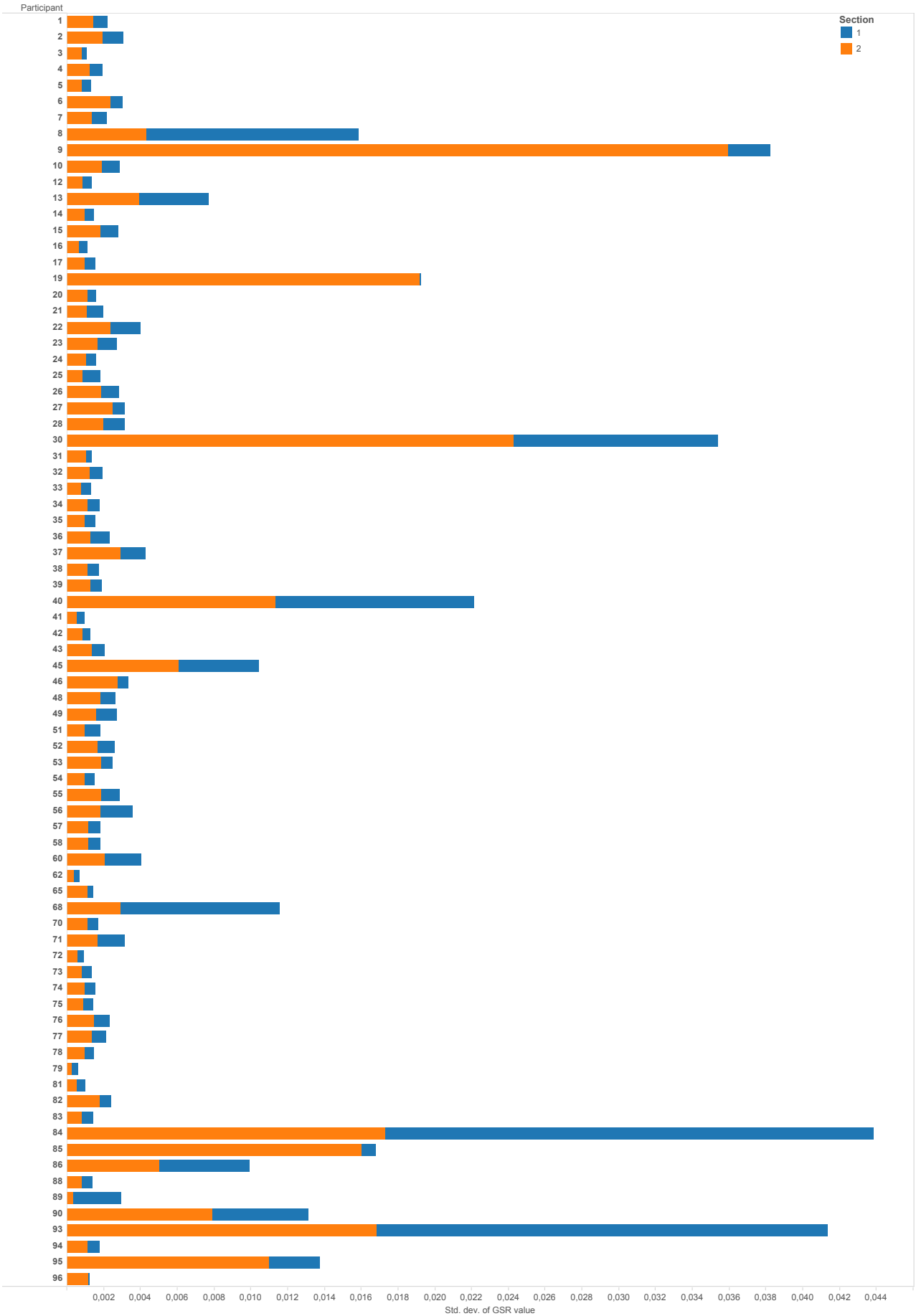

Figure 10: Within-subject variability in Sections 1 and 2.

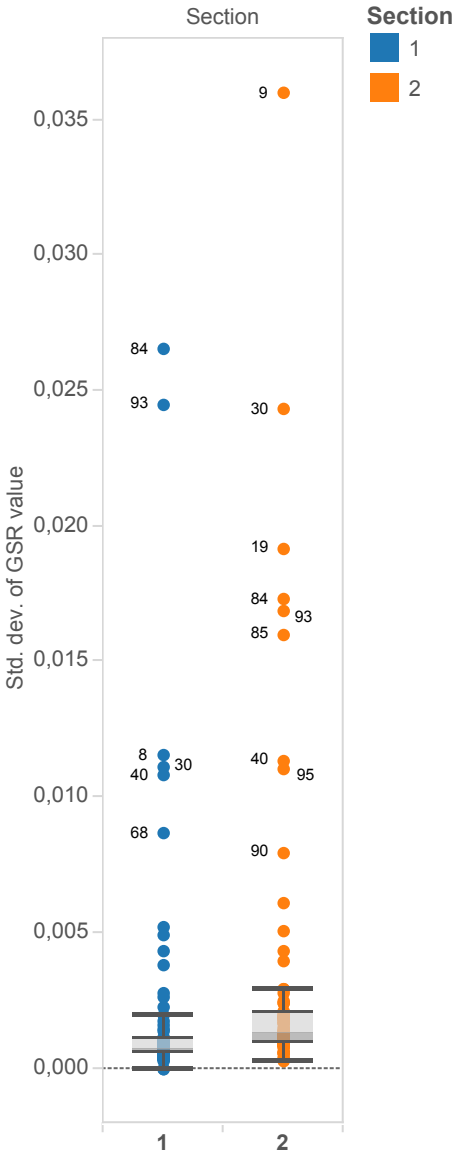

Figure 11: Inter-quartile range and outliers in Sections 1 and 2, with details shown for participant.

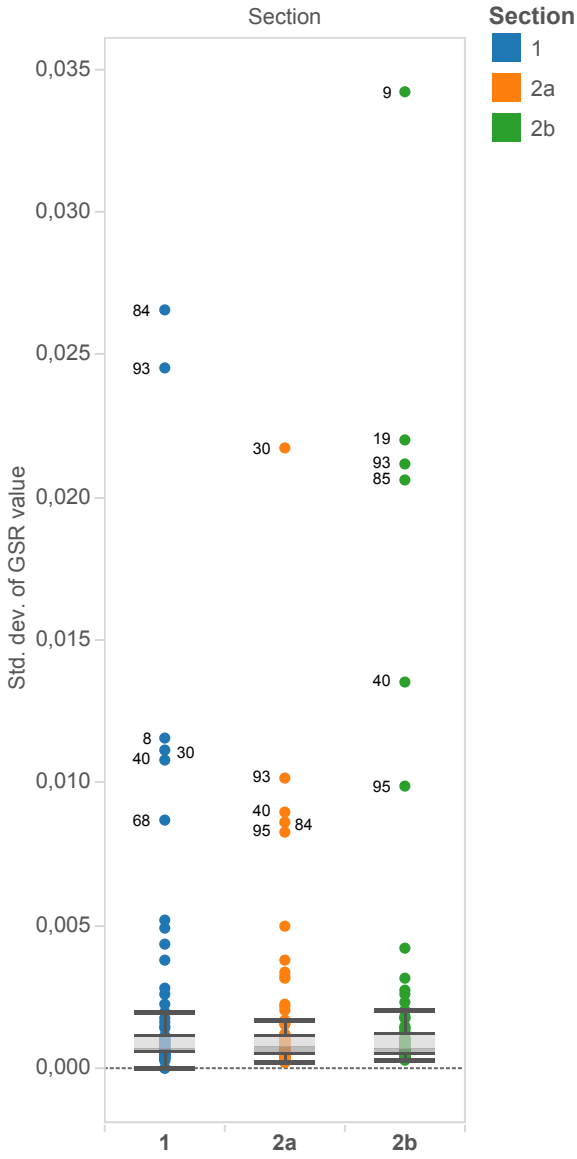

Figure 12: Inter-quartile range and outliers in Section 1 and Subsections 2a and 2b, with details shown for participant.

and when compared with Figure 11 it is evident that the increased affective arousal in Section 2 is the consequence of subsectional differences, however small they may seem.

Discussion

In this paper we have chosen to use ET and EDA measures as examples of popular, simple and relatively inexpensive tools of data col-

lection for experimental audience-testing research, which can be combined with textual grammar analysis of visuals and sound as suggested by Kress and van Leeuwen (2006) and van Leeuwen (1999). In this way, we wish to highlight the fruitful synergies that can be potentially realised by the inter-paradigmatic approach. We believe that the modality markers make up a powerful toolbox for describing systematically the parametric details in the construction of aesthetic objects, which can be related subsequently into the attentional and affective aspects of aesthetic experience. On the one hand, results of textual grammar analysis – being valuable in their own right – can support the generation of hypotheses to be tested empirically, against which the analysis results can be further discussed and put into perspective. Both confirmation and refutation of predictions made by the textual grammar analysis will be interesting and can inform future strategic use of aesthetic elements. On the other hand, grammar-analysis results could support experimental testing, by providing a more informed basis for selecting stimuli in relation to a specific study purpose. For instance, an analysis might lead to the result that two or more photos are not sufficiently comparable, given that the parametric differences are too many to draw revealing conclusions; and that for a real study it might be necessary to collaborate with a cook and a photographer to create a set of images that are sufficiently alike to isolate the effect of the independent variable (veggie vs. non-veggie). The same applies to moving images, such as TV commercials. However, in the case of the tyre commercial chosen for the present paper in which we made internal comparisons between narrative-based sections (and subsections), a text-oriented close reading, as the one outlined above, seems, under any circumstances, to be a necessary prerequisite for understanding differences in affective response.

Now, by focusing on ET and EDA, concerning respectively the attentional and affective aspects of an aesthetic experience, thus leaving out the *cognitive* aspect (cf. Marković 2012), the present paper by no means pretends or claims to be exhaustive. Considering viewers' self-reported verbal responses to aesthetic objects, making sense of such responses, and the use of findings for strategic-communication purposes, may indeed represent a significant complementary component of the audience-testing part. For the same reason, we emphasize once again the positioning-stating intent of the

paper. In our view, the synergies are not limited to the specific tools and methods used in this paper. Indeed, other systemic-functional and social-semiotic approaches to grammar analysis and the wide range of different computing techniques for measuring audiences' attentional, cognitive and affective response all have interesting synergies that can be fruitfully explored to obtain new and interesting results in the study of aesthetic experience of art.

Acknowledgments

We would like to acknowledge the students at Applied Aesthetics for coming up with the idea of comparing the two kinds of food in an eye-tracking experiment and for their participation in the experiment. We also wish to thank The Eye Tribe for generous sponsorship of trackers.

References

Bergstrom, Jennifer Romano and Andrew Schall. 2014. *Eye Tracking in User Experience Design*. Elsevier Science.

Bernstein, Basil. 1971. *Class, Code and Control: Volume 1. Theoretical Studies Towards a Sociology of Language*. London: Routledge.

Bernstein, Basil. 1981. "Codes, Modalities, and the Process of Cultural Reproduction: A Model." *Language in Society* 10(3): 327–363.

Bojko, Aga. 2013. *Eye Tracking the User Experience: A Practical Guide to Research*. New York: Rosenfeld.

Boucsein, Wolfram. 2012. *Electrodermal Activity*. New York: Springer.

Edwards, Anna. 2013. "The Most Terrifying Advert You Will Ever Watch! Japanese Tyre Commercial that Comes with Its Own Health Warning (So Don't Blame Us…)." Accessed February 7. http://www.dailymail.co.uk/news/article-2517390/Terrifying-Japanese-tyre-advert-Commercial-comes-health-warning.html

Green, Shannon Rae. 2013. "Warning: Scariest Ad Ever." Accessed February 7. http://www.usatoday.com/story/news/world/2013/12/04/japanese-ad-scary-tire-autoway/3871073/

Halliday, Michael Alexander Kirkwood. 1978. *Language as Social Semiotic*. London: Arnold.

Halliday, Michael Alexander Kirkwood. 1985. *An Introduction to Functional Grammar*. London: Arnold.

Huey, Edmund B. 1898. "Preliminary Experiments in the Physiology and Psychology of Reading." *American Journal of Psychology*, 9(4): 575–886.

Junk, Julian. 2011. "Method Parallelization and Method Triangulation: Method Combinations in the Analysis of Humanitarian Interventions." *German Policy Studies* 7(3): 83–116.

Kraut, Alan G., and Daniel W. Smothergill. 1978. "A Two-Factor Theory of Stimulus-Repetition Effects." *Journal of Experimental Psychology* 4(1): 191–197.

Kress, Gunther, and Theo van Leeuwen. 2006. *Reading Images: The Grammar of Visual Design*. London: Routledge.

Kress, Gunther. 2010. *Multimodality: A Social Semiotic Approach to Contemporary Communication*. London: Routledge.

Land, Michael F. 2014. *The Eye: A Very Short Introduction*. Oxford: Oxford University Press.

Lauwereyns, Jan. 2012. *Brain and the Gaze: On the Active Boundaries of Vision*. Cambridge, MA: The MIT Press.

Magnussen, Rikke, Nikita Kharlamov, Maria Zachariassen and Birger Larsen. 2016. "Knowledge Generation in Technology-Enhanced Health Exhibitions: Using Eye-Tracking Methods to Understand Audience Knowledge Generation in Health Promotion Exhibitions." In *Proceedings of the 15th European Conference on e-Learning ECEL*, edited by Jarmila Novotná and Antonín Jančařík, 448–460. Charles University, Prague.

Marković, Slobodan. 2012. Components of Aesthetic Experience: Aesthetic Fascination, Aesthetic Appraisal, and Aesthetic Emotion. *i-Perception* 3(1): 1–17.

Monnet, Agnieszka Soltysik. 2015. "Body Genres, Night Vision and the Female Monster: *REC* and the Contemporary Horror Film." In *Monstrous Media/Spectral Subjects: Imaging Gothic from the Nineteenth Century to the Present*, edited by Fred Botting and Catherine Spooner, 143–156. Manchester: Manchester University Press.

Morris, Charles W. 1939. "Esthetic and the Theory of Signs." *Journal of Unified Science* 8(1–3): 131–150.

O'Toole, Michael. 1990. A Systemic-Functional Semiotics of Art. *Semiotica* 82(3-4): 185–209.

Ravaja, Niklas. 2004. "Contributions of Psychophysiology to Media Research: Review and Recommendations." *Media Psychology*, 6(2): 193–235.

Rudner, Richard. 1951. "On Semiotic Aesthetics." *Journal of Aesthetics and Art Criticism* 10(1): 67–77.
Russell, James A. 1980. "A Circumplex Model of Affect." *Journal of Personality and Social Psychology*, 39(6): 1161–1178.
Sobchack, Vivian. 2004. *Carnal Thoughts: Embodiment and Moving Image Culture*. Berkeley, CA: University of California Press.
Uno, Tadao, and William W. Grings. 1965. "Autonomic Components of Orienting Behavior." *Psychophysiology* 1(4): 311–321.
Van Leeuwen, Theo. 1999. *Speech, Music, Sound*. London: Macmillan Palgrave.
Wheelwell, Donnya. 2000. "Against the Reduction of Art to Galvanic Skin Response." *Journal of Consciousness Studies*, 7(8–9): 37–42.
Withnall, Adam. 2013. "Is This the Scariest Advert Ever? Japanese Tyre Commercial Comes with Its Own Disclaimer and Health Warning." Accessed February 7. http://www.independent.co.uk/news/world/asia/is-this-the-scariest-advert-ever-japanese-tyre-commercial-comes-with-its-own-disclaimer-and-health-8980781.html

Note

1 The ninth marker, movement, constitutes an addendum to the former eight markers when it comes to moving images. According to Kress and van Leeuwen, "movement can be represented with different degrees of realism or abstraction and hence play a role in modality judgements", just "[l]ike visual detail, background, depth, light and shade, colour etc." (2006, 264).
2 As for the domain of sound, van Leeuwen distinguishes between naturalistic, sensory and abstract-sensory coding orientation (1999. 177–180).
3 http://theeyetribe.com/.
4 http://www.ogama.net/. Ogama version 5.0 was used.
5 The TV commercial can be accessed online via YouTube: https://www.youtube.com/watch?v=54U6BgYuJMY.
6 See Sobchack (2004) for a cognitive-phenomenological study on the immersive qualities of "subjectively perceived and embodied presence" in first person point-of-view cinematic images (2004, 136).
7 For the impact of stimulus repetition, see, e.g., Uno & Grings (1965) and Kraut & Smothergill (1978).