Emotional Response to the Audio-visual Pattern Language of Film

Gesche Joost - Deutsche Telekom Laboratories, Berlin, Germany, (+49) 175 292 62 99, gesche.joost@telekom.de
Noam Tractinsky -Ben Gurion University, Beer Sheva, Israel (+972) 8 647-2226
Sandra Buchmüller - Deutsche Telekom Laboratories, Berlin, Germany, sandra.buchmueller@telekom.de, (+49) 30 8353-58366
Roman Englert - DTAG Laboratories at Ben Gurion University, Beer Sheva, Israel, roman.englert@telekom.de, (+49) 171 520-5755

Abstract

In this paper, we evaluate a notation system for Audio-visual Patterns (AVP) that visualizes auditory and visual parameters of film. This system is the basis for a comparative analysis of the rhetorical structure and for predicting the emotional response of viewers of film. We conducted two experiments to test the prediction of the proposed AVP system by asking female viewers about their emotional responses to 10 selected TV commercials. The results indicate that the AVP system provides sound predictions regarding the arousal level of the viewers as the commercial unfolds.

Conference theme: Design & Emotion: Methodological issues

Keywords: Pattern language, film analysis, rhetoric, emotion, persuasion, multimedia, arousal, design research

Introduction

Film as an audio-visual medium has a high potential to arouse the emotions of the audience. Throughout the past decades, film makers developed a variety of techniques to elicit intense emotions (Mikunda 1986, Scheuermann 2008) and therefore improved their ability to seduce the audience (Bonsiepe 1961). How these emotions can be predicted, measured and evaluated is a challenging research topic that is currently addressed by various disciplines (Fogg 2003, Hagemann et al. 1999, Rottenberg et al. 2007, Smith 1996). In this paper, we examine a visualization technique that attempts to predict emotional arousal. This idea goes back to Gui Bonsiepe's approach on "Audio-Visual Rhetoric" (Bonsiepe 2008) where he suggests to use diagrams instead of mere verbal description to discover and analyze audio-visual micro-structures in media. These micro-structures can be described as audio-visual patterns (AVPs). In this paper we briefly describe a notation system for AVPs. A fuller exposition of the system is presented in (Joost 2008). The main focus of the paper is in demonstrating, by means of two empirical studies, the viability the system and its ability to predict the audience's emotions.

The Theory of AVP

Using the AVP method leads to a notation protocol of the rhetorical structure of film. We learn from rhetorical scholarship that certain AVPs refer to a specific level of arousal (Joost et al. 2008). According to their kind and quantity, they indicate a higher or lower emotional impact on the audience: high emotional arousal is often derived from a larger number of AVPs and patterns like hyperbole, climax or acceleration; lower emotional impact is derived from using only few AVPs such as illustrative music or variation. Therefore, the protocol can be used to predict the emotional response of the viewers, because it indicates the level of emotional arousal at each point in time as being low, medium, or high (the corresponding reactions to the three types of rhetoric styles: *logos*, *ethos* or *pathos*). The research question we follow in this paper is whether the predictions derived from this visualization of AVPs match the actual emotional response of the audience.

Notation Practice

The structure and style of a film can be visualized in notation protocols (see, for example, Figure 1 and 2). These protocols are described in more detail in (Joost et al. 2008). The development of tension (which we propose is predictive of the audience's level of arousal) is indicated by a curve, which depends on the structure of the AVPs. Thus, in Fig. 1 the horizontal dimension shows the development of the film in time, whereas the vertical line shows the visual patterns and the expected emotional intensity. For each film clip, some special points of interest

can be marked, called emotional check points (ECPs). These ECPs mark specific AVPs within the commercial that are likely to affect the audience's level of arousal. For example, one can notice a sharp increase in expected arousal between ECP1 and ECP2 in Figure 1. In Figure 2, we see an overall low level of arousal throughout the commercial. Therefore, we expect a much higher emotional arousal after watching the commercial in example 1 than after watching example 2.

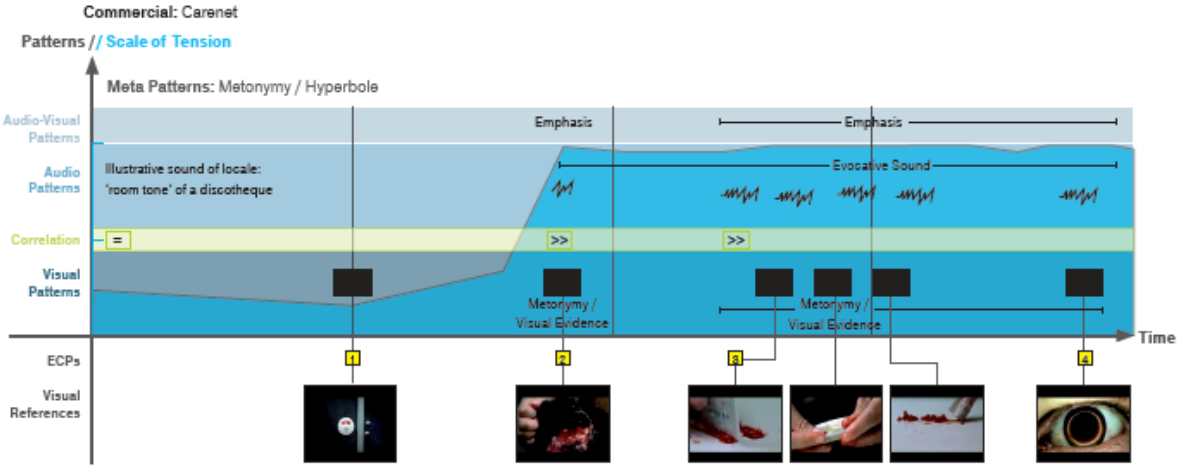

Figure 1. Notation protocol of the commercial "Carenet"

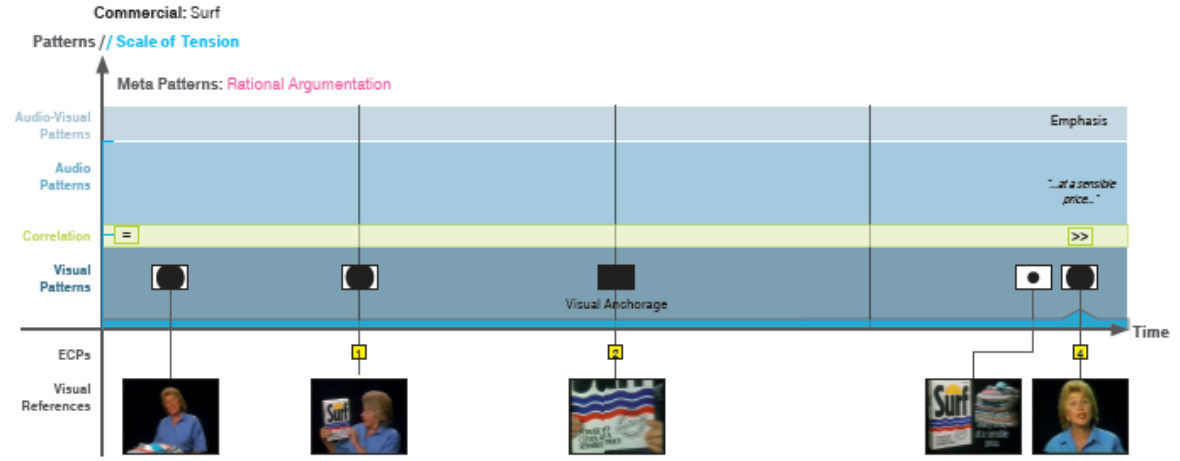

Figure 2. Notation protocol of the commercial "Surf"

Empirical Assessment

We conducted two experiments to gather empirical evidence regarding the efficacy of the AVP method. For the experiment, we selected a set of 10 TV commercials varying in the level of potential emotional arousal from high to low. For each commercial, the main rhetorical patterns where defined in the auditory and visual channel. While the commercials were selected by

German researchers, who also defined their patterns and predicted their emotional outcome, the empirical evaluation was conducted in Israel. This allowed us to test the universality of the predictions that result from the AVP method.

The video clips

The 10 commercials were selected from public internet databases like YouTube. The commercials can be accessed online at http://www.geschejoost.org/notations-protokolle. The commercials ranged in length from 27 to 61 seconds, with an average of 41 seconds. To minimize the potential confounding effect of content on observed emotions, we selected clips with minimal amount of speech. Thus, the evaluation was based mostly on visual style and sound. We tried to select clips that varied in terms of the potential level of arousal, although this turned out to be difficult: nowadays, most of the commercials have high emotional impact (ethos or pathos). It is quite difficult to find commercials relying on rational argumentation (logos). The commercial "Surf detergent" (Figure 2), which was selected as targeting the lowest level of arousal, was, in fact, produced in the 1980's. For each clip, two of the authors, who are also designers, prepared a pattern protocol using the AVP method (e.g. Figure 1 and 2). The protocols of the ten clips were then compared to the emotional reactions of the participants in the two experiments, as described below.

Classification of expected emotional reactions

To demonstrate the viability of the AVP notation system in predicting emotions of the viewers, we use the Emotional-Rhetorical Response Model (see Figure 3). On the horizontal scale it marks emotions ranging between "pleasant" and "unpleasant" and on the vertical scale it indicates the emotional arousal from "sleepy" to "aroused".

This demarcation follows the conventions in research on emotions [Russel 2003]. To this two-dimensional model of emotions we mapped the rhetorical model that differentiates between three levels of arousal (*logos*, *ethos* and *pathos* as described above). The rhetorical model only partially overlaps the emotional dimensions. Because rhetoric only focuses on communicative purposes fostering positive values of emotional arousal, it only covers the range between "neutral" to "aroused", but not the rest of the arousal dimension. This theoretical argument is supported by everyday observations of audio-visual segments that were created to deliver a message: it is hard to find such segments that on purpose attempt to deactivate the audience.

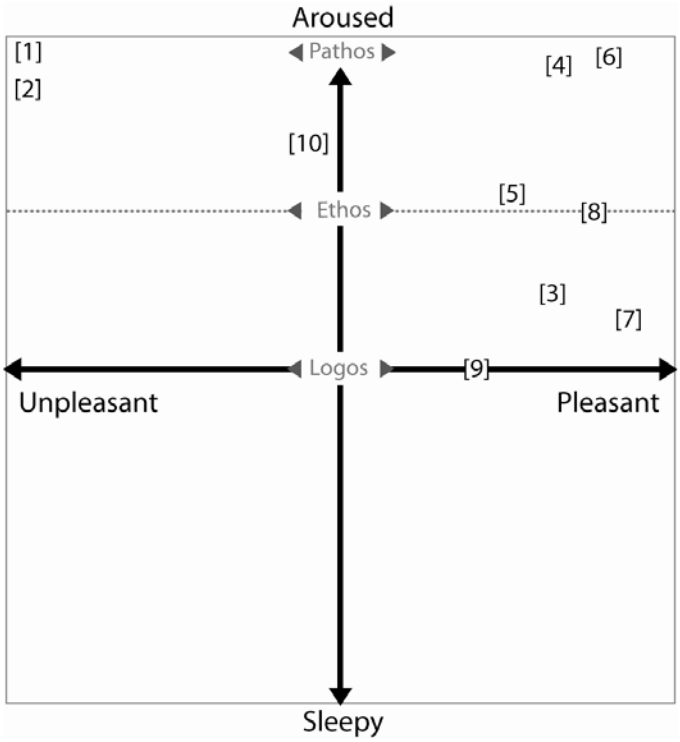

Figure 3. Expected levels of arousal and pleasure for 10 commercials; expected values were distributed in the emotional space subjectively by the designers based on AVPs

As discussed before, audio-visual stimuli such as film segments, video clips, commercials, and animations can be categorized according to their level of arousal on the basis of the pattern structure. However, from a rhetorical viewpoint, the pattern structure provides only scarce information regarding the degree of pleasure or displeasure (e.g., *ethos* suggests moderate levels of pleasance). Therefore, there is a strong subjective component in predicting the audience reactions along the "pleasure" dimension.

Based on AVPs, the two authors responsible for the classification indicated the expected audience's reaction on the Emotional-Rhetorical Response Model. From this process we learned that some patterns are not strictly linked to only one level of arousal. Thus, it seems that there is a more rough matching of AVPs to the level of arousal [Joost et al 2008] referring to categories like higher or lower emotional impact. Rather, it is the combination of quantity of patterns, level of tension and the actual type of AVP that provides basis for estimating the emotional effect it will have on the audience. For example, a high amount of patterns with higher emotional impact like hyperbole, climax, or antithesis is likely to produce a high level of arousal. Next, we describe the two experiments that were designed to empirically evaluate the AVP-based classification of the video clips.

Experiment 1

Objective

The objective of experiment 1 was to evaluate whether the designers' predictions of the emotions evoked by the AVPs match the actual emotions experienced by viewers of the commercials. It is worthwhile to note that the differences between the designers and the viewers in the experiments (who did not have any background in design) included also national differences as the designers were German and the participants were Israelis.

Sample

Ten participants who responded to advertising on campus bulletin boards in an Israeli university volunteered to participate in the experiment. Each participant received 100 NIS (about $24) as a compensation for her participation. Because of the exploratory nature of this study we preferred to increase the likelihood that the participants indeed express their emotions. Thus, we chose only females as participants based on the findings that, relative to men, women tend to exhibit stronger reports of emotional experience (Hagemann et al. 1999, Rottenberg et al. 2007).

Stimuli

The participants evaluated the 10 video clips described above. Post-experimental interview verified that none of the participants had seen any of the commercials before the experiment.

Procedure

The experiment was administered individually in an office room by the same experimenter in two sessions, which were conducted a few days apart within the same week. In the first session, the participants received instructions regarding the experimental procedure and signed a consent form. They then viewed five of the clips one at a time. After each clip, they were asked to fill out a questionnaire. The questionnaire included items describing both general emotions (i.e., arousal and pleasure) as well as discrete emotions (e.g., love, fear). The response to each item was made on a 9-point scale. This procedure was repeated in the next session for the other 5 clips. Because of space limitations we report here only on the results that pertain to general emotions.

Results

Overall Clip evaluations

The average ratings of the emotions that the participants experienced right after viewing the commercials are presented in Table 1. For each of the general emotions, the data in Table 1 has

endpoints of -4 (very unpleasant, very sleepy) and +4 (very pleasant, very arousing). We transformed all of the arousal scores by a positive linear adjustment of 0.7. The adjustment was made to bring the average observed arousal evaluations to the same baseline of the expected baseline, i.e., a score of 0 for the least arousing clip (Clip 9).

Clip	Subject	Pleasure	Arousal
1	Anti Child Abuse	-3.4	3.6
2	Carenet	-3.1	3.5
3	Panty Liner	0.9	1.6
4	Lays Chips	1.6	3.6
5	Lufthansa NewYork	0.3	2.7
6	Mercedes S Class Hamster	0.5	2.5
7	Colgate Toothbrush	1.9	1.4
8	Eva Longoria for Loreal	2.1	1.7
9	Surf detergent	1.0	0.0
10	CWS Washroom Solutions	0.3	3.1

Table 1. Average observed arousal and pleasure scores

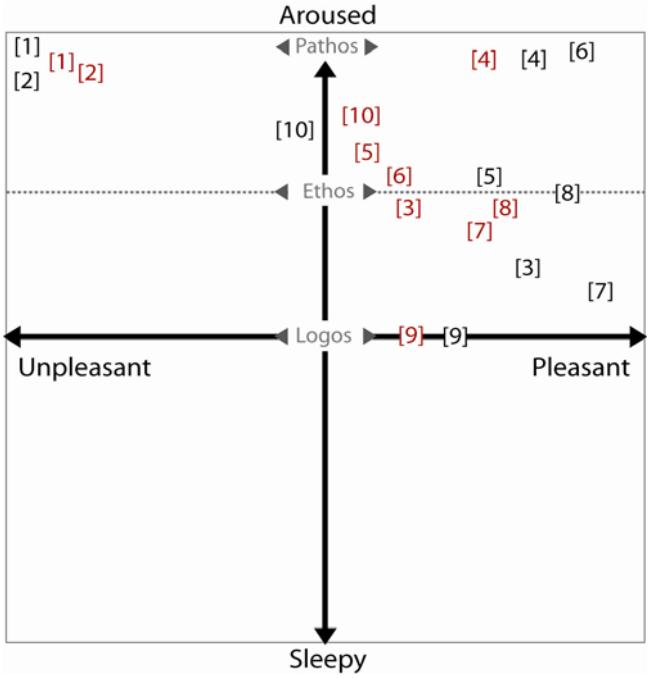

Figure 4. Expected (black) vs. observed (red) levels of arousal and pleasure for 10 commercials; expected values were distributed in the emotional space by the designers based on the AVPs and their subjective evaluation; observed values are the empirical results of Experiment 1 with endpoints at -4 to +4 for both emotions.

Figure 4 displays the average arousal and pleasure scores of each of the clips. The clips are identified by their serial number (from 1 to 10).

Each clip is represented in Figure 4 by two marks, reflecting the clip's position on the emotions grid: (A) the expected (hypothesized) level of emotions generated by the clip (denoted by a black number). (B) The observed level of emotions, which were calculated as the average evaluations of the 10 participants after they viewed the clip (denoted by a red number). Because the expected clip position on the grid is a rough and subjective estimate produced by the two designers based on the AVPs, it is impossible to conduct statistical tests regarding the deviations of the observed emotions from the expected emotions. Still, we can make some inference from the data. It is clear that the observed ratings of six of the clips are very similar to the designers' expectations (Clips 1, 2, 4, 8, 9 and 10). The actual ratings of the other four clips are more distant from the designer's expectations. However, the observed arousal levels of three of those 4 clips (3, 5, and 7) are quite similar to the expected levels. It seems, though, that the designers have overestimated the degree of pleasantness of these four clips.

Ranking-based analysis

The above analysis compared the expected and the observed absolute values of arousal and pleasure for each clip. However, it should be noted that assigning such values is quite arbitrary and may present biases due to different evaluation criteria of the designers and of the viewers or due to various differences in culture and in response tendencies. Perhaps a better way, at this stage of evaluating the AVP's approach, is to assess the relative difference (or similarity) between how designers' judged the distribution of the clips on the emotional grid and the actual distribution that was derived by the participants' assessments. Table 2 shows the order according to which the designers expected the clips to rank on the pleasure and on the arousal continua and the observed (i.e., averaged over 10 participants) ranking. This information is also depicted in Figures 5 and 6. Spearman's rank correlations between expected and observed rankings were .82 for Pleasure ($p=.01$) and .87 for Arousal ($p<.01$), suggesting very high correspondence between the designers' judgments and the participants' experience.

	Pleasure		Arousal	
Rank Order	**Expected**	**Observed**	**Expected**	**Observed**
10	Clip 7	Clip 8	Clip 1	Clip 1 (tie)
9	Clip 8 (tie)	Clip 7	Clip 6	Clip 4 (tie)
8	Clip 6 (tie)	Clip 4	Clip 4	Clip 2
7	Clip 4	Clip 9	Clip 2	Clip 3
6	Clip 10 (tie)	Clip 10	Clip 3	Clip 5
5	Clip 5	Clip 6	Clip 5	Clip 6
4	Clip 9	Clip 5 (tie)	Clip 8	Clip 8
3	Clip 3	Clip 3 (tie)	Clip 10	Clip 10
2	Clip 2	Clip 2	Clip 7	Clip 7
1	Clip 1	Clip 1	Clip 9	Clip 9

Table 2: Rank order of expected and observed scores on the Pleasure and Arousal dimensions (1 = lowest, 10 = highest).

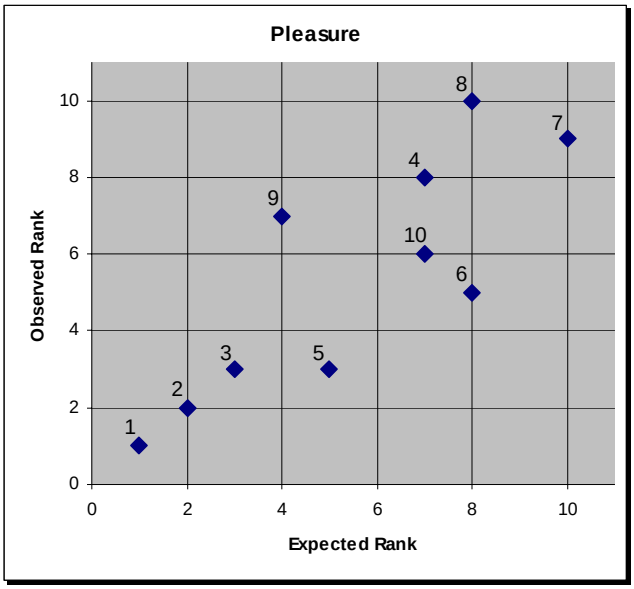

Figure 5: Rank order of expected and observed Pleasure rankings of the 10 clips. The numbers above the data points in the graph identify the video clips.

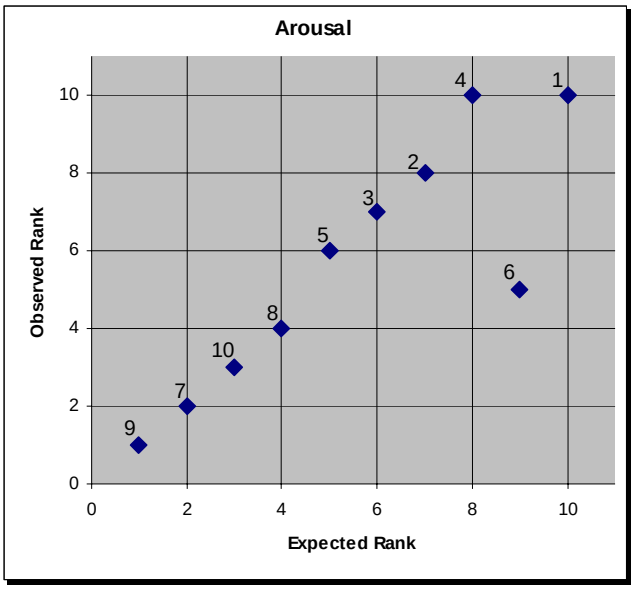

Figure 6. Rank order of expected and observed arousal rankings of the 10 clips. The numbers above the data points in the graph identify the video clips.

Experiment 2: Evaluations of clip segments

Objective

The objective of this study was to test, *within commercials*, whether viewers would experience the same changes in arousal as predicted by the designers based on the AVP system.

Sample.

Ten females, who did not participate in the previous study, responded to advertising on campus bulletin boards and volunteered to participate in the experiment. Each participant received 50NIS (about $12) as a compensation for her participation.

Stimuli

Five of the video clips from Experiment 1 were used in this experiment. The criterion for selecting the 5 commercials was that, based on the AVPs, each of those commercials was predicted to contain noticeable differences in expected arousal levels between some of its segments, which were defined by ECPs. Each commercial was then divided into 3 to 5 individual segments that were presented to the participants and evaluated as described below. Clip lengths varied from 29 seconds for Clip 4 to 61 seconds for Clip 10. The average length was 52 seconds. Like in Experiment 1, post-experimental interview verified that none of the participants had seen any of the commercials before the experiment.

Procedure

The procedure was similar to that of Experiment 1 with the following difference. To evaluate the participants' emotions during watching the commercials we used the retrospective think

aloud (RTA) method [Guan et al. 2006, Van den Haak et al. 2004]. Immediately after viewing each of the five video clips, the participants watched the clip again, segment by segment. At the end of each segment (that is, at the ECP), they were asked to reconstruct their general feelings and discrete emotions while they watched the commercial for the first time. Again, due to space limitations we report here only on the participants' rating of their arousal level by the end of the segment, which is the most direct and important aspect of emotions that the AVP method attempts to capture.

Results

Table 3 presents the expected and the observed changes in arousal from one segment to the next within each commercial. The expected changes are based on the changes in the AVPs from segment to segment at the ECP (see Figure 1 for an example of changes in tension/arousal between Segments 1 and 2 in Clip 2). The observed changes are the average ratings of the 10 participants in the study of their arousal level at the end of each segment. The observed arousal ratings were compared using repeated measures analysis of variance that included the arousal ratings of two subsequent segments as its dependent variables.

Clip No. (1)	Compared Segments (2)	Expected Arousal Change (3)	Observed Arousal Change (4)	Expectation Supported? (5)
1	2 vs. 1	Major increase	Increase ***	+
	3 vs. 2	No change	ns	+
	4 vs. 3	Very minor decrease	ns	(+)
	5 vs. 4	Minor increase	Decrease **	-
2	2 vs.1	Major increase	Increase ***	+
	3 vs. 2	No change	ns	+
	4 vs. 3	No change	ns	+
4	2 vs. 1	Major increase	Increase **	+
	3 vs. 2	No change	ns	+
	4 vs. 3	Decrease	Decrease *	+
5	2 vs.1	Decrease	Decrease *	+
	3 vs. 2	Very minor increase	ns	(+)
10	2 vs. 1	Increase	ns	-
	3 vs. 2	Decrease	ns	-
	4 vs. 3	Increase	Increase **	+
	5 vs. 4	No change	Decrease *	-

Table 3:Comparison of expected vs. observed changes in arousal level in subsequent clip segments. Observed changes were determined by repeated measures ANOVA.

Thus, for example, the first line of data in Table 3 reflects the following facts. (1) A comparison is made within Clip 1. (2) The comparison is made between the first and the second segments. (3) The AVPs indicate a substantial *expected* increase in level of arousal from the end of Segment 1 to the end of Segment 2 (note that this prediction is based on the consensus estimate of the two designers). (4) The *observed* arousal ratings also indicate an increase in arousal from Segment 1 to Segment 2, and this difference is statistically significant at the .01 level. (5) The observed ratings match the designer's expectations. Overall, of the 16 segment comparisons in this study, in 10 the subjective reports were entirely commensurate with the AVP-based expectations. In 2 cases, expectations for minor arousal changes were not matched by statistically significant changes. In four cases an expected change was not accompanied by a comparable empirical change.

Discussion

The AVP method can serve developers of audio-visual media in various ways. In this study, we examined the predictive power of the method in anticipating the emotions of viewers of TV

commercials. The predictions were based on the combination of quantity of audio-visual patterns, the level of tension and the type of the AVPs in the commercial. In terms of the emotions that the audio-visual material intends to generate, the method is geared primarily to describing the induced arousal level, and to a lesser extent the degree of induced pleasance. To test the predictive ability of the method we applied the AVP method to ten TV commercials and later observed users' reactions to those commercials. The first experiment demonstrated a good fit between the predicted arousal level at the end of the clip and the observed arousal level of the study's participants. Although the AVP method was not designed to provide a quantitative measure of the experienced emotion, the average arousal ratings of six of the ten clips were very close to its predictions, and three other clips were quite close as well. This indicates, perhaps, that further work can improve the model's quantifiability and predictive ability. A more appropriate test, at this stage, of the adequacy of the model is a rank-based comparison between predicted and observed levels of emotion. Indeed, the correlation between the observed and the predicted order of the clips on the arousal scale was strong and provided support for the model's validity. Still, it is interesting to examine the three cases in which there was some discrepancy between the predicted and the observed levels of emotions. In particular, the discrepancy was high in Clip 6, which was a commercial for the Mercedes S Class. The participants in our study commented that they had a hard time understanding the clip and its message and relating to it in any way. This may explain the fact that instead of the predicted strong feeling of arousal and pleasance, the participants, on average, felt quite neutral about the clip: mildly aroused and somewhat pleased. This highlights a potential weakness in the model, namely that it does not account for the content of the message. Thus, if the audio-visual cues are not aligned with the content the result may be confusing for the audience and their emotions may deviate from those predicted by the AVP model. Furthermore, this result indicates that the persuasion of the audience – the basic rhetorical goal of commercials – was not successful in this case. This could be due to cultural differences (the commercial could only aim to address an audience with German cultural background, not Israeli cultural background), or to the commercial's producers overestimating the visual and auditory effect of the clip. The reason for this mismatch cannot be identified based on the AVP method, but the incongruity between rhetorical structure and emotional reactions of the audience indicates that the intended persuasion went wrong.

The correspondence between the predicted and the observed pleasance of the clips was not as strong as that of the arousal level. This was anticipated, as the AVPs are only marginally intended to predict pleasure. Still, the rank-order correlation between predicted and observed pleasance was high. We found that viewers' ratings of the clips on the pleasant-unpleasant scale were more conservative than the expected ratings, especially for the positive half of the scale. This finding suggests that it may be more difficult for multimedia designers to induce pleasing emotions relative to inducing arousal in the viewers.

The second experiment examined whether the AVPs are sensitive enough to allow prediction of emotional changes (along the arousal dimension) within video clips. This is not a trivial objective, as the examined clips were quite short (the average clip length was 52 seconds). The results again demonstrated the usefulness of the method in predicting emotional changes. If we ignore 2 very minor predicted changes, which were not detected by the users, then 10 of 14 within-clip predicted changes (or no-changes) in arousal were confirmed empirically.
Overall, the results support the proposition that the AVP method can serve, among other things, as a sensitive and useful tool to predict the emotional influence of multimedia stimuli. The fact that the predicted emotions were done by German designers whereas the ratings of actual emotions were provided by Israeli viewers contributes to the apparent validity of the method. Still, we consider this study as a first step in the assessment of this method. It is obvious, as we found out in this study, that the method has its limitations as well. It deals with structural aspects of the stimulus, but not with its complement – the content. Thus, the developers need to balance and synchronize both aspects of the multimedia application. Further work has to be done also on the library of AVPs, e.g. an investigation of interactive media like games. For this context, the research question is whether different media might use different sets of patterns. With this further work, the library of AVPs can be stocked up step by step to become a powerful tool for Human Computer Interaction (HCI), too. With respect to the automating of the process, the authors are currently examining the possibility to retrieve auditory patterns in media based on algorithms. These patterns could be mapped to standardized AVPs and their potential of emotional arousal. With this approach, one can envision an automatic process for predicting emotional response to audio-visual media.

Conclusion

This paper tackles the problem of analyzing and predicting the emotional effects of audio-visual media on the recipient. We evaluated a system of audio-visual patterns that visualizes the dynamic structure of film segments and predicts the emotional reactions of the audience to those segments. As such, the system can serve as a meaningful aid for designers of film and media applications. We conducted two experiments to test the proposition that the system can be useful in predicting the emotional states (especially arousal) of viewers of multimedia messages. The results attest to the usefulness of the AVP system as a method for dealing with the research problem.

References

Anderson, L., Shimamura, A. (2005). “Influences of emotion on context memory while viewing film clips”. The American journal of psychology 118, 323-337.

Bonsiepe, G. (2008). „Audiovisualistische Rhetorik in zeitbasierten Medien: Über die kognitive Relevanz diagrammatischer Visualisierungen“. Joost, G., Scheuermann, A. (Eds.) Design als Rhetorik, Birkhäuser Verlag Basel, 231-246.

Bonsiepe, G. (1961). “Persuasive Communication: Towards a Visual Rhetoric”. Crosby, T. (Eds.) Uppercase Nr. 5, London, UK, 19–34.

Fogg, B.J. (2003). “Motivating, Influencing, and Persuading Users”. Jacko, J.A. and Sears, A. The Human-Computer Interaction Handbook: Fundamentals, Evolving Technologies, and Emerging Applications. Lawrence Erlbaum Associates: Mahwah, NJ.

Guan, Z., Lee, S., Cuddihy, E., Ramey, J. (2006). “The Validity of the Stimulated Retrospective Think-Aloud Method as Measured by Eye Tracking”. Proc. CHI 2006 ACM Press, 1253-1262.

Hagemann, D., Naumann E., Maier, S., Becker, G., Lürken, A., Bartussek, D. (1999). „The assessment of affective reactivity using films: Validity, reliability and sex differences”. Personality and Individual Differences, 26, 627-639.

Joost, G. (2008). „Film-Sprache. Die audio-visuelle Rhetorik des Films“. Bielefeld.

Joost, G., Buchmüller, S., Englert, R. (2008). „Audio-visual Rhetoric: Visualizing the Pattern Language of Film”. Proceedings of Design Research Society Conference, Sheffield, UK.

Mikunda, C. (1986). “Kino spüren: Strategien der emotionalen Filmgestaltung”. München, C.H. Beck.

Rottenberg, J., Ray, R.R., Gross, J.J. (2007). “Emotion elicitation using films“. J.A. Coan, J.A., Allen, J.J.B (Eds.) „The handbook of emotion elicitation and assessment”. Oxford University Press, New York, USA.

Russell, J. A. (2003). “Core affect and the psychological construction of emotion”. Psychological Review 110, 1, pp. 145-172.

Scheuermann, Arne (2008). „Affekttechniken des Design“. In Joost, G., Scheuermann, A. (Eds.) Design als Rhetorik, Birkhäuser Verlag Basel, 221-228.

Smith, M. (1996). “Emotions and the Structure of Narrative Film: Film as an Emotion Machine”. Lawrence Erlbaum Associates, Mahwah NJ, USA.

Van den Haak, M.J., de Jong, M.D.T., Schellens, P.J. (2004). “Employing think-aloud protocols and constructive interaction to test the usability of online library catalogues: a methodological comparison”. Interacting with Computers, 16, pp. 1153-1170.