Profiling Arousal in Response to Complex Stimuli using Biosignals

Felix Putze, Dominic Heger, Markus Müller, Christian Waldkirch, Yves Chassein, Ivana Kajic and Tanja Schultz

Institute of Anthropomatics, Cognitive Systems Lab, Karlsruhe Institute of Technology (KIT), Karlsruhe, Germany

Keywords: Affective Computing, Biosignal-based Arousal Profiling, Person-independent Classification.

Abstract: We investigate the use of biosignals (blood volume pressure and electrodermal activity) for person-independent profiling of arousal responses to complex, long-term stimuli. We report the design of a user study with 14 subjects to elicitate affective responses with films of different genres. We present a detailed analysis of the recorded signals and show that it is possible extract information on the differences between films and within each film from biosignals. We use this information to automatically discriminate four film classes in a person-independent fashion with an accuracy of up to 97.8%.

1 INTRODUCTION

In human interaction, humans sense a large variety of signals of the persons we are interacting with and try to infere information about their cognitive and affective mental state. The term affective computing was coined by Rosalind Picard (Picard, 2000) to describe computer systems which can estimate and react to the user's affective state. This state can for example be assessed based on biophysiological signals continuously emitted by the human body. Today, researchers widely agree that the perception of a human state in a particular situation is of central interest for intelligent systems to enhance system performance to achieve a better user experience. Since the beginning of affective computing, a large number of studies and systems has been published on the estimation of affect from biosignals ((Picard et al., 2001), (Lichtenstein et al., 2008), (Soleymani et al., 2008)). Arousal is one of the most important aspects of affect. It is related to information evaluation and task performance. This paper contributes to this area a detailed analysis of arousal profiles (i.e. the degree of arousal or its correlates over time) as response to longer-lasting, complex stimuli in form of complete films of different genres. In HCI, this is highly relevant as most interaction sessions, especially in the entertainment domain, consist of a long sequence of interacting affective stimuli. We investigate the possibility of extracting information on the arousal profile from biosignals and using this information for automatic discrimination of films and film segments.

2 EXPERIMENTAL SETUP

To collect data of multiple persons covering different long-lasting dynamic affective states, we designed an experiment using full short films for affect elicitation. During the presentation, we recorded physiological responses to those films. For our study, we selected three films of different genres: The first one is a zombie horror film with socially critical undertone (AKUMI). It starts with a slow exposition, than introduces horror elements and culminates into a showdown battle. The second film is a slow, silent stop-motion arthouse-film with a constantly low suspense curve (FLOWERS). The third one is a humorous animation film riddled with slapstick jokes (LIFTED) about alien driving school. Before each film, we included a relaxation phase (RELAX) of approximately two minutes to bring participants back to a neutral, calm state. A relaxation phase consisted of a sequence of nature stills with meditative music. To counter ordering effects, we randomly assigned participants to two different permutations of the films. This ensures that for every pair of two films, there are recordings for both possible orderings. After each film, participants filled out a film related questionnaire to classify its genre and to indicate their emotional response to it using a Self Assessment Manikin (SAM, (Bradley and Lang, 1994)). During each film or relaxation phase, we recorded EDA and blood volume pressure using plethysmography (PPG). We showed the films in a dimmed, windowless and empty room to avoid as much distraction as possible. Participants sat in a fix-

Putze F., Heger D., Müller M., Waldkirch C., Chassein Y., Kajic I. and Schultz T..
Profiling Arousal in Response to Complex Stimuli using Biosignals.
DOI: 10.5220/0004249203470350
In *Proceedings of the International Conference on Bio-inspired Systems and Signal Processing* (BIOSIGNALS-2013), pages 347-350
ISBN: 978-989-8565-36-5

a fixed position in front of a large projection screen on which we showed the films to create an intense impression. Biosignals were recorded using a wireless biosignal monitor by PLUX[1] with a sampling rate of 1000Hz. In total, we recorded about 23 minutes of biosignals for each participant. 14 participants attended our study. All of them were students or visitors at the KIT between 14 and 30 years. Four participants were female. We ensured by analyzing content-related questionnaires that participants perceived the three presented films as sufficiently different.

For self-assessment of emotions experienced during the film, a subscale of the SAM technique is used to measure arousal. Arousal is comparable for the engaging films AKUMI and LIFTED (3.29 and 3.23) and significantly lower for the slow paced FLOWERS (2.53). As we expected people to experience a variety of emotional states during the course of each film, we combined the SAM-technique with a time axis: After watching each film, participants state their arousal continously over course of the whole film by drawing a curve from the beginning of the film to the end. To give orientation for the participants, we presented stills of the film from every minute as reminders of the film flow. For analysis, we sample each curve at every full minute to derive a self-assessed arousal profile. The average range of arousal values over the course of a film is as high as 1.72, which is higher than the differences between averages for different films. This indicates that even a short film with a clear affective tendency (e.g. a horror film) has a dramaturgy that results in both low and high arousal values and which should also be reflected in the biosignal recordings. This also indicates that we cannot simply use an average arousal score to label the complete corresponding biosignal stream but that we have to take a more detailed look at the dramaturgic structure.

3 SIGNAL ANALYSIS

In the next step, we analyze the variation of the recorded biosignals in relation to the dramaturgic structure of a film. Starting with an overall comparison of the different films, we see a significant difference ($p < 0.05$) in mean EDA signal amplitude between FLOWERS and AKUMI and between FLOWERS and LIFTED. However, as we already saw that the arousal profile to a film varies strongly over time, we now look at temporal patterns in the recorded biosignals. In order to trace the affect changes that were elicited through different types of films and within each film, we look at the EDA signal because a change in skin conductivity occurs quickly as a response to an increased level of arousal and can be interpreted in the time domain.

As each individual EDA curve contains a lot of session specific effects which cannot easily be attributed to events in the film, we generate an averaged EDA curve for each film from the data of all participants. For each participant, the EDA signal is normalized and lowpass filtered at 0.5 Hz. Afterwards, we calculate the averaged EDA signal. It is correspondingly scaled and illustrated in Figure 1 for the films AKUMI and FLOWERS[2]. When we compare both curves, we notice a number of differences: The averaged EDA signal for AKUMI shows strong temporal variation while the signal for FLOWERS is relatively smooth. This is caused by a larger number of sharp rises of the EDA signal for AKUMI as a respose to unexpected, surprising or exciting events (Ekman et al., 1985) in the individual signals. We call these rises *startles*. The lack of startles for FLOWERS also causes the monotonic decreasing trend of the curve. Taking the connection between arousal and EDA activity into account, we can draw the conclusion that certain events in the horror film AKUMI in general cause a higher arousal than the slowly paced arthouse film FLOWERS. We also see that the averaged EDA signals roughly match the trend of the averaged discretized arousal curves we extracted from the SAM (with a time delay of ca. 1 minute).

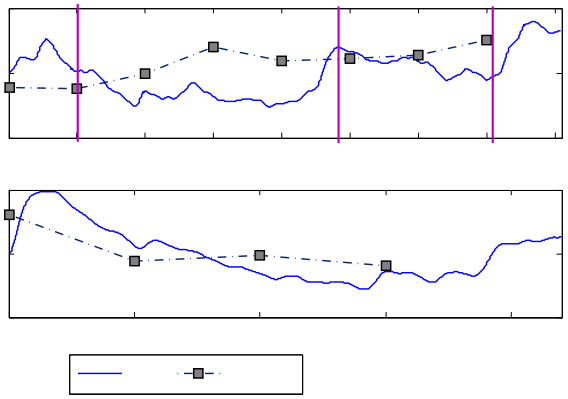

Figure 1: Top: Averaged EDA curve for the film AKUMI with indices for marked scenes; Bottom: Averaged EDA curve for the film FLOWERS.

Still, the EDA signal for AKUMI is not stationary across the whole film. Can we determine which effects cause the observed peaks and level changes? Comparing the EDA signal to the arousal curve, we

[1]http://www.plux.info

[2]The plots are generated by averaging individually z-normalized EDA curves and are therefore unit-free.

only see a very rough match. However, when reviewing the film material directly, we get a more precise picture: We can identify certain scenes which drastically change the dramaturgy of the film. These scenes are marked by vertical lines in Figure 1. The strong EDA rises at the second and third marker correspond to climatic scenes of the film while the valley between seconds 60 and 240 corresponds to a slow-paced and monolog-driven part of the film. We can quantify this effect by automatically measuring the tempo of the film as number of scene changes in a sliding time window. The resulting graph in Figure 2 (top) shows that a rising tempo around second 240 corresponds very well to the lasting rise of the EDA curve and indicates a strong general affective response to this change in dramaturgy. This type of analysis is not possible for FLOWERS because of the employed stop-motion technique. To explain the course of the EDA curve for this film, we instead compare it to one generated from the corresponding RELAX sections (see Figure 2 (bottom)) and see a very similar monotonous trend in the signal after a short rise at the start of each phase. Only the begin of the final credits (which are missing for RELAX) marks the onset of a slow rise. This indicates that the arousal profile for the slow paced FLOWERS is comparable to the one for RELAX. In summary, we see significant differences in the EDA-derived arousal profiles in response to films of different genre and style and also to different sections within one film.

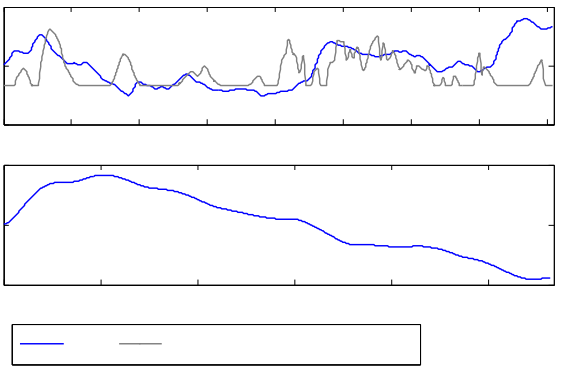

Figure 2: Top: Comparison of EDA curve with number of scene changes for AKUMI; Bottom: Averaged EDA curve for a RELAX phase.

4 MOVIE CLASSIFICATION

We will now use those observed characteristic, biosignal-based arousal profiles as features for automatic classification of films and film segments. We start by investigating the classification of the recorded biosignals into four classes: the three different films as one class each and all relaxation phases combined as another one. We design a person-independent setup to investigate the ability of differentiating films based on their arousal profiles.

The first step of the classification process is a z-normalization of the raw data for each subject to level different signal baselines. We segment the data into windows of 20 seconds lengths with an overlap of 10 seconds. On each window, we extract tentative features: For the PPG signal, we extract the mean heart rate and its variation by applying peak detection on the bandpass filtered signal. For the EDA signal, the number of startles is determined automatically using peak detection after low-pass filtering. We then extract the mean EDA value and the number of startles. We now derive the final features which describe the complete arousal profile of one film based on the whole signal stream. Those features consist of statistics on the window-based features for each film: mean, maximum and mimimum value, standard deviation, relative peak position, peak width and relative centroid. Final features are independent of the film length to avoid leakage of this information to the classification process. This results in a feature vector of 28 dimensions. After feature extraction, we train and evaluate a Naive Bayes classifier using leave-one-person-out cross-validation. Within each iteration of the cross-validation, we use Sequential Forward Feature Selection (SFFS) on the training set of the respective fold to select the best features. To evaluate features during this selection process, an inner cross-validation within each fold is performed using stratified sampling. Using this setup, we achieve a very high average accuracy of 97.8% with a minimum precision of 88.24% and a minimum recall of 93.3% over all classes. The standard deviation between folds is 8.3%, which indicates relative stability in the face of small changes of test and training data.

To investigate the generalization abilities and the feature stability, we calculated a histogram on the selected features. On average, the feature selection resulted in a feature set of size 2.45 (median: 2). The left part of Table 1 gives an overview of the most frequently selected features over all cross-validation folds. It indicates that there are some stable features which are regularly picked over others; Only 8 of 28 features are ever selected. This result indicates that those features generalize well across persons. When training a model using the features from Table 1 instead of using SFFS for each fold, we still achieve an average recognition rate of 95.56%, indicating that the original recognition accuracy was not the result of over-specialization. To investigate the predic-

Table 1: Left: Most frequently selected features over all cross-validation folds. Table shows the selection frequency in percent for features derived from mean heart rate (MHR), variance of heart rate (VHR), mean EDA (EDA) and number of startles (STA): Peak Width (PW), Relative Peak Index (RPI), Average (AVG), Standard Deviation (SD), Maximum (Max), Minimum (Min). Right: Same information for feature selection without peak width, peak position and centroid.

full feature set			restricted feature set		
Signal	**Feat.**	**Freq.**	**Signal**	**Feat.**	**Freq.**
MHR	PW	53.3	EDA	SD	100.0
EDA	PW	46.7	MHR	Avg	73.3
MHR	SD	33.3	MHR	SD	73.3
MHR	Avg	13.3	MSC	Avg	66.7
MHR	Min	13.3	VHR	Min	40.0
STA	PW	13.3	EDA	Min	33.3
STA	SD	6.7	STA	Avg	33.3
VHR	RPI	6.7	MHR	Max	26.7
			MSC	Max	26.7
			VHR	Avg	26.7

Table 2: Average accuracies in percent for pairwise classification of windows from different minutes (1 = 0s to 60s, 2 = 60s to 120s, ...) of the film AKUMI.

Min.	**2**	**3**	**4**	**5**	**6**	**7**	**8**
1	61	54	67	64	75	70	56
2		52	58	59	72	70	67
3			64	63	77	73	63
4				58	72	70	66
5					70	67	61
6						59	67
7							58

tive power of both modalities separately, we see similar recognition rates if we restrict the feature set to only EDA-based features (97.8%) or only PPG-based features (96.67%). We conclude that both modalities carry information on the arousal profile and depending on the application it may be possible to reduce the number of required sensors.

Note that some of the employed features (e.g. relative centroid position) encode information specific to the dramaturgy of the films. Therefore, the trained model will not be applicable to different films without loss of recognition accuracy (albeit, films of similar dramaturgical structure could work). We therefore repeat evaluation after removing the features encoding relative peak position, relative centroid position and peak width. As expected, the accuracy drops significantly to 73.3%. The merit of this model is that it still provides reasonable recognition accuracy using much more generic features which promise generalizability to different films. The selected features are given in the right part of Table 1. Again, we identify a number of features which is repeatedly selected across folds.

As documented in Section 3, significant differences cannot only be noted between different films but also during the course of one film. To investigate the possibility of identifying different parts of the film based on biosignals, we classify the window-based features extracted for the process described above for the movie AKUMI. To each window, we assign a label based on its position within the film, using one label for each full minute. Classification is performed pairwise for each combination of two labels to investigate similarity effects. In this setup, we do not expect high classification accuracy for each pair of segments. Instead, we can interprete the recognition accuracy as a measure of distance between two segments based on the arousal profile. Table 2 presents the results of leave-one-person-out cross-validation. As expected, performance reaches levels of up to 77% for sections which are dramaturgically very different. For sections which are similar in this regard (e.g. both from the fast-paced ending), accuracy drops. This result is in strong accordance with the observations on Figure 1 and shows that, even given the difficulty induced by fuzzy class transitions, automatic affective profiling of a film is possible.

REFERENCES

Bradley, M. and Lang, P. (1994). Measuring emotion: the self-assessment manikin and the semantic differential. *Journal of behavior therapy and experimental psychiatry*, 25(1):49–59.

Ekman, P., Friesen, W., and Simons, R. (1985). Is the startle reaction an emotion?. *Journal of Personality and Social Psychology*, 49(5):1416.

Lichtenstein, A., Oehme, A., Kupschick, S., and Jrgensohn, T. (2008). Comparing two emotion models for deriving affective states from physiological data. In Peter, C. and Beale, R., editors, *Affect and Emotion in Human-Computer Interaction*, volume 4868 of *LNCS*, pages 35–50. Springer Berlin / Heidelberg.

Picard, R. (2000). *Affective computing*. The MIT press.

Picard, R., Vyzas, E., and Healey, J. (2001). Toward machine emotional intelligence: Analysis of affective physiological state. *Transactions on pattern analysis and machine intelligence*, pages 1175–1191.

Soleymani, M., Chanel, G., Kierkels, J. J. M., and Pun, T. (2008). Affective characterization of movie scenes based on multimedia content analysis and user's physiological emotional responses. *Multimedia, International Symposium on*, 0:228–235.