Towards Detecting Cognitive Load and Emotions in Usability Studies Using the RealEYES Framework

Randolf Schultz, Christian Peter, Michael Blech, Jörg Voskamp, and Bodo Urban

Fraunhofer Institut für Graphische Datenverarbeitung Rostock, Joachim-Jungius-Straße 11, 18059 Rostock, Germany
{randolf.schultz,christian.peter,michael.blech,joerg.voskamp, bodo.urban}@igd-r.fraunhofer.de

Abstract. In this paper, we will discuss some extensions to the RealEYES framework that can help to automatically detect interesting sections in usability studies using additional sensor input and knowledge discovery techniques.

Keywords: usability, emotions, cognitive load, human performance monitoring.

1 Introduction

Usability test systems usually collect a huge amount of various data. Screen recording, gaze tracking, mouse and keyboard input are just basic components of data streams nowadays usability test systems generate. Further data sources like audio data and face monitors add to the ever growing wealth of data to be processed and even newer technologies like emotion detection and human performance monitoring emerge and promise to add to the quality and value of usability studies.

Processing and analysing those data is very expensive in time and human resources, even if good tools for visualisation and analysis of the data are available. This is because common tools cannot find interesting sections in the test, where the subject experienced high mental load or prominent emotions, automatically. Instead, the usability expert has to browse manually to supposedly critical positions in the test data for analysis. Consequently, critical spots not envisioned by the expert may be missed.

The goal of the work presented here is to apply automatic analysis algorithms for identifying ongoing emotions and high cognitive load in the user to speed up the analysis process and spot critical situations in the data stream more easily. Commercially available tools do not offer equivalent functionality and are not extensible in a way we would need it. Consequently, we will discuss actual extensions to the RealEYES framework that help to automatically detect critical situations in usability studies by use of novel sensors and knowledge discovery techniques.

This paper is organized in the following way: first, the original RealEYES framework will be introduced briefly, followed by a description of the extensions for detecting emotions and cognitive load. We close with a discussion of exemplary study results and give directions for further work.

N. Aykin (Ed.): Usability and Internationalization, Part I, HCII 2007, LNCS 4559, pp. 412 – 421, 2007.

2 The RealEYES Framework

The RealEYES framework combines a number of tools to support the entire process of a usability study, from preparation, execution, and analysis to communicating test setup data, measurement data and test results in an efficient manner.

The previous version of the framework already supported a multitude of data types in its data backbone: meta data on the test, screen-shot videos and video capture of the user, audio data, gaze and mouse pointer positions, and application and test specific events. The data is collected and synchronized in a single datastream per subject. After recording, the test data of the subjects may be visualized or analyzed in manifold ways including playback and statistical analyses.

2.1 The Structure of the Framework

The RealEYES framework consists of the following components: Recorder, Analyzer, Statistics, and Questioning (see Fig. 1).

They all work on a common information backbone comprised of AVI and XML files. The AVI files contain video and audio streams as well as gaze, mouse, keyboard, and event data, and the XML files contain meta data and intermediate test results. Thus, no lengthy conversion of data is needed when working with the tools in the RealEYES framework. All tools are equipped with sophisticated export functions for an easy integration of the test results into a usability report.

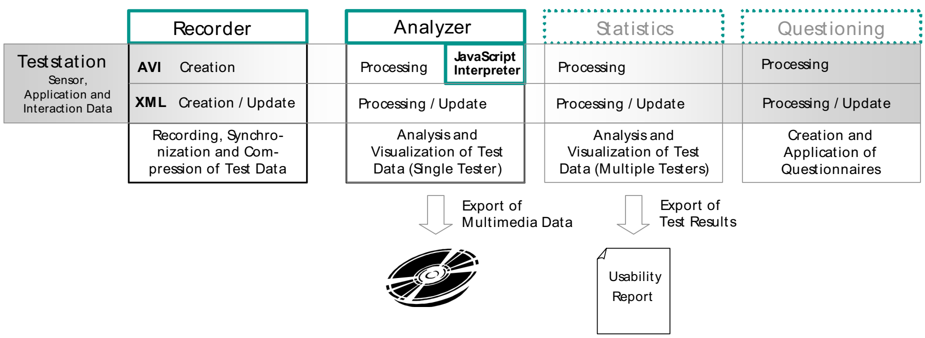

Fig. 1. Structure of the RealEYES framework

2.2 The Tools of the Framework

The most important tools in the RealEYES framework are Recorder and Analyzer. The *Recorder* manages the recording of all data. It requests the meta data, captures, synchronizes and compresses the test data and writes it to a single AVI file per session. The *Analyzer* is the main analyzing application that replays all video streams and visualizes the other data. Many standard and advanced visualizations of the screen-shot video together with the gaze and mouse data are available, such as a temperature grid; compare Fig. 4. The adjustable visualizations offer detailed insights into the actual user interactions with the tested product. New visualizations may be implemented by an advanced user of the system and plugged into the Analyzer.

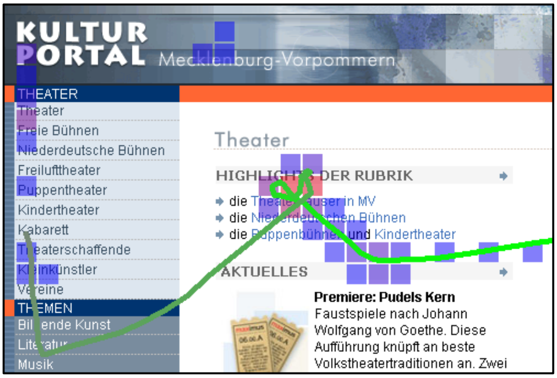

Fig. 2. Temperature Grid Visualization

The Analyzer is a tool for offline analysis of the collected test data and supports navigation using standard video player controls as well as event marks (see lower part of analyser window shown in Fig. 2). Event marks may be inserted automatically by the Recorder, manually by the usability expert during the test and the analysis phase, or semi-automatically using scripts written in JavaScript that run in the Analyzer's script engine during offline analysis. Regions of interest may be defined and simple statistical calculations can be performed on them. The regions may not only be defined geometrically but can also be bound in time or to certain tasks. Other tools of the framework may resort to these regions.

To improve the ability to test web-applications that often do not fit completely on the screen, both Recorder and Analyzer support and obey a scrolling region in all their features (see Fig. 4).

The two other tools in the RealEYES framework are Statistics and Questioning. The *Statistics* tool allows for complex statistical analyses to be accomplished on the acquired data (see Fig. 7). Furthermore, the Statistics tool is able to analyze and visualize data from all sessions of a study, i.e. to perform analyses over data of different users of a study. To illustrate, the Statistics tool can answer questions like "Did the majority of the subjects see the navigation buttons" or "What's the average time users looked at the advertisement". The *Questioning* tool allows the usability expert to create and utilize online questionnaires. The data gathered from the questionnaires is written to the XML file and can be processed by the Statistics tool.

3 Extensions of the RealEYES Framework

To further improve the frameworks capabilities, new technologies have been tested. Since user experience is closely coupled with emotions and cognitive load, sensors measuring related physiological parameters have been tried and finally incorporated into the framework.

For reliably detecting emotional states, various modalities can be used ([13], [7]). While face data of the user are already available in the data as well as speech, analyzing them is still a challenge (cf. [4]), particularly when it comes to facial feature analysis of subject thinking aloud. Progress has been made in recent years particularly in speech analysis for emotional and cognitive signs (see also [4], [15]) and also in facial feature extraction ([5]; [1]).

The modality most researched and best understood today in terms of emotion is physiology, so we decided to add physiology sensors as well. After having examined various commercial sensor systems and decided against them for their inappropriate sensing elements and wires, we opted for the EREC sensor system developed by Fraunhofer IGD Rostock (see [11] and [12]). The RealEYES framework has been extended to support the open EREC protocol and Analyzer and Statistics tools have been extended to visualize and analyze the additional data.

3.1 Sensor Hardware

The additional sensors to gather physiological data should be minimally intrusive and easy to use. The EREC system (cf. [11]) developed at Fraunhofer IGD Rostock is a first step towards this. It features sensors in a fingerless glove, not hindering human computer interaction, built-in reliability checks, and wireless data transfer. The system has been tested in several studies and has been continuously improved (see [12]). Currently, heart rate, galvanic skin response and skin temperature are measured and made available immediately to the recording tool of the RealEYES framework. The sensor system is equipped with special error detection and reliability checks, which makes inclusion of physiological data into the processing application fairly easy. Particularly, EREC's convenient way of providing the data in engineering units made it very easy to incorporate the data without the need of implementing proprietary conversion algorithms. Also, with EREC implementing the SEVA standard for self-validating sensors ([6]; [3]), sensible data (from the processing point of view) are always provided, equipped with a standardized reliability flag for fast and easy appraisal of the data (see Fig. 3). For more details on the EREC system, refer to [11] and [12].

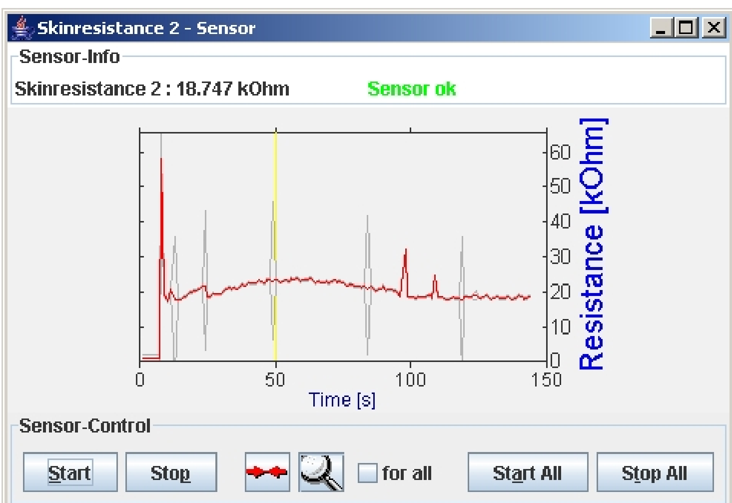

Fig. 3. EREC data are provided with sensor state and reliability information. The double spike zones indicate spots of uncertainty. Note that valid data are available for processing anyway. The line at 50ms represents an event mark.

3.2 Emotion Detection on Physiology Data

Sensing emotion-related information delivers a huge amount of data, with an enormous quantity of parameters being extracted from it, cf. [2]. At the current state of research, it seems to be necessary to collect physiological data at a relatively high sample rate. From our experience, 20 measurements per second are a good choice. This leads to 1200 samples per device and minute, accumulating to 36000 measured values per device for a half-hour experiment. To process the data, we apply knowledge discovery techniques because they allow to deal with big data sets and to examine them without previous specification of hypotheses or parameters to use. The extracted information can then be used to define attributes, statistical methods, learn algorithms, and classification concepts for integration into emotion detection classifiers, cf. [14].

Corpus. Based on a corpus previously build at Fraunhofer IGD Rostock (see [8]), filter operations and classifiers for emotional and cognitive states have been developed. They can be used to derive emotional states and cognitive overload from physiological data (heart rate, skin resistance and temperature). Those classifiers are to be integrated into the analyzing tools of the RealEYES framework.

In addition to the recording of physiological data, speech records are suitable to recognize current emotional states, as well. The research in emotion recognition from speech shows that the extraction of acoustic and prosodic features in combination with machine learning techniques leads to robust classifiers with success ranges from 60-80%, see [9].

To get more experience, we started an own study with an already existing speech corpus which is taken from the speech database of the TU-Berlin (cf. [10]). From this corpus we took 500 samples from male and female actors who have spoken several times a set of simple sentences, each time with an emotional intention from seven different categories.

In a first step, filters were applied that extract basic acoustic features such as duration, pitch, intensity and frequencies. In the following step, statistical calculations were applied on these basic features to characterize the current sample. Finally, the combination of the acoustic features with common statistical variables (min, max, median,…) and further typical speech processing values (longest voice, relation speech to non speech,…) leads to about 70 valuable features. The summary of these statistic features over all 500 samples is the basis for the final machine learning process. After performing tests with several different classifiers, we have reached a prediction performance of 76% with a support vector machine (compare Table 1).

The gained know-how for emotion detection from speech will be used in our next studies to improve the evaluation and interpretation of the study results. To be sure, to get valuable speech from a study, a test person is invited to give comments while performing a test. In addition, our speech corpus will be extended with the speech samples from new studies, so that the prediction performance and robustness of our speech classifiers can increase with every new study.

Table 1. Prediction performance on physiological and speech corpus

Physio.	*Best Classification Result*	*Random Classifier*
Max	49% (Euphoric)	20%
Min	26% (Helpless)	20%
Typical	38%	20%
Speech		
Typical	76%	14%

Store Study Results. To deal with the huge amount of data resulting from different studies, we have implemented a database to store emotion-related multimodal data, from physiological or speech recording modalities. The database scheme is designed to hold data from a (theoretically) unlimited number of sessions, of different studies, using any combination of input modalities. With this database, studies can be compared to other studies and data can be classified more precisely.

3.3 Integration into RealEYES Tools

To integrate the aforementioned sensing and emotion detection technologies into the RealEYES tools we decided to exploit the various extension methods already provided by the framework. Those are visualizations in the Analyzer and Statistics tool and events in the Analyzer.

Display of Raw Physiological Data. We started out with a graph display of the raw physiological data in the Analyzer (compare Figure 4). Using the graph display of the raw data, rapid changes in the physiological data (e.g. skin resistance) hinting stress may be discovered easily by the usability expert. The screen space available for this visualization, unfortunately, is not high enough to give a good overview of the data acquired during a complete test session. However, a rapid change is also easy enough to compute with JavaScript, leading to a script for the semi-automatic insertion of event marks. The event marks calculated by the script cover a complete test session and make it easy to navigate to all supposedly critical sections in the test. Nevertheless, we also plan to integrate a visualization of the raw data in graph form also in the Statistics tool. Here we could calculate and display values for a complete session and even for all test subjects from all sessions of the test at once. It would also be possible to create event marks in our data streams for exploitation in the Analyzer. This could be done even for situations that expose no conspicousness in the data of a single test subject.

Display of Classification Results. A visualization of classification results was already created in the context of our EmoTetris study (see [8]). We displayed a comic face to visualize a predominant detected base emotion and used a star plot diagram to visualize the results of all classifiers for base emotions (see Figure 5).

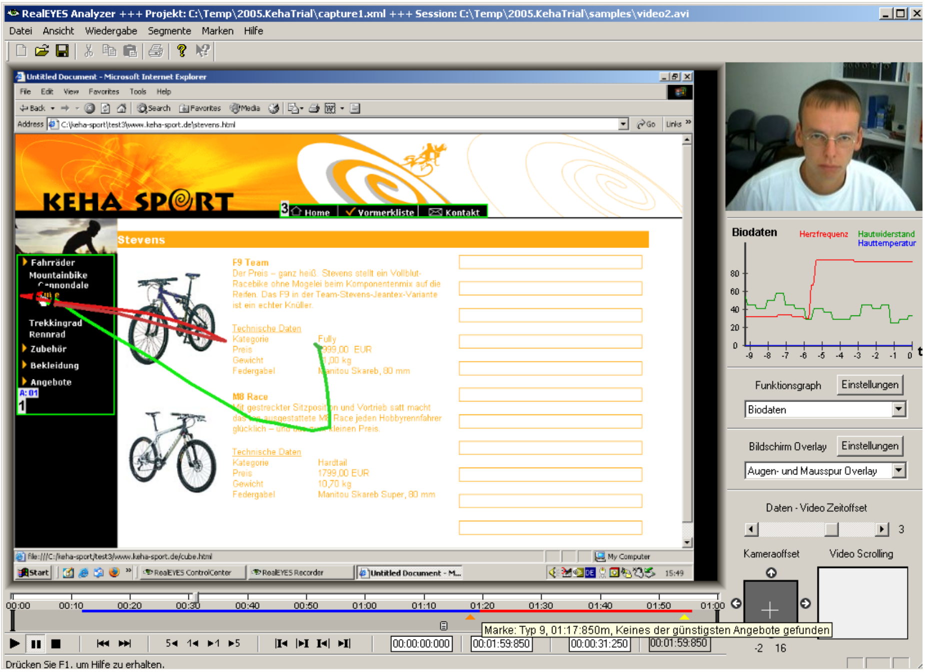

Fig. 4. RealEYES Analyzer Tool with Visualization of raw physiological data (below the image of the test subject) and Event Marks (below the time line)

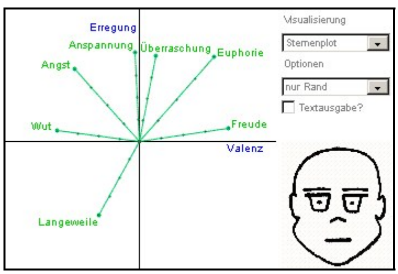

Fig. 5. Visualization of Emotion Classification Results in EmoTetris; Star Plot of Base Emotions in Russel Diagram (*left*), Comic Face (*lower right*)

Since this visualization only shows a current state and no process, it is not very well usable in the RealEYES context and we plan to integrate a graph display similar to the graph display for raw physiological data into the Analyzer and the Statistics tools instead. Similar to the script that detects rapid changes in physiological data, event marks will be inserted when a classifier detects e.g. a high stress level.

Measures must be taken to not flood the data stream with events, especially when classifiers for different modalities are active. Further problems arise when the classifiers deliver contradicting results. Those problems may be solved by creating a second classifying layer that works on the output of the classifiers for different modalities.

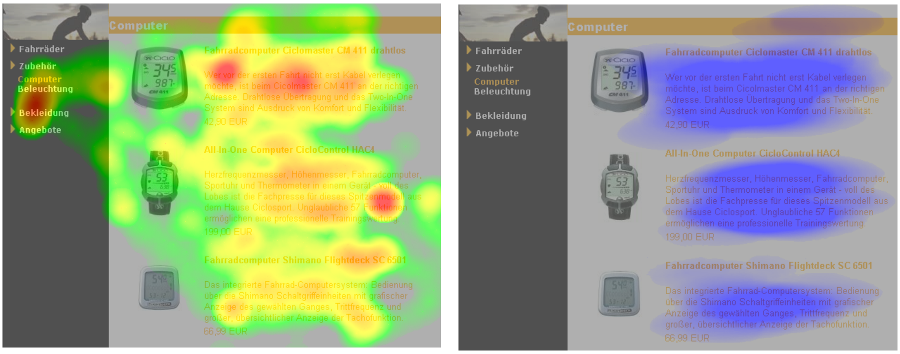

Fig. 6. Conventional Heat-Map (*left*) and Read-Map (*right*) of two test subjects in comparison

In our EmoTetris study we discovered that apart from the basic emotions, states like loss of control also play an important role in the human computer interaction. Other states that are relevant in the context of usability are searching and reading. Consequently, we developed classifiers for those states and integrated them in the Statistics tool. The classifiers work solely on gaze and mouse data as input. The output of the classifier may easily be mapped to the screen-content of the recorded application guided again by the gaze data. The result is an enhanced version of a Heat-Map, visualizing parts on a screen image that have been read (rather than just looked at) by the subject (see also Figures 6 and 7), hence the name Read-Map. While the left image of Figure 6 only shows that the test subjects looked at the text sections on the underlying web page, the right image makes it clear that the texts actually have been red. For usability studies, Read-Maps give valuable hints for actual reading activity (an interesting section of a usability test) on e.g. E-Learning content, or advertisements.

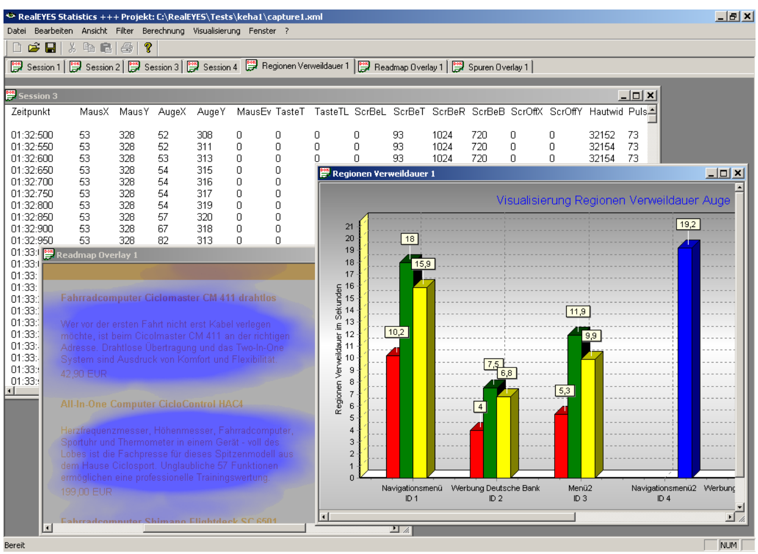

Fig. 7. User Interface of RealEYES Statistics with ReadMaps on the text display

4 Conclusions

In this paper we discussed some extensions to the RealEYES framework that allow to integrate and process sensor data with machine learning algorithms to automatically find critical incidents in usability studies. We presented some detection results, showing that our approach is not only applicable but opens new perspectives in usability studies. Further work will focus on improving the detection rates using more physiological parameters (breathing activity, oxygen level) and improved knowledge discovery algorithms. We would also like to integrate another in-house-developed tracking system for human motions to extend our scope beyond the desktop scenario.

References

1. Aleksic, P.S., Katsaggelos, A.K.: Automatic Facial Expression Recognition Using Facial Animation Parameters And Multi-Stream Hmms. In: IEEE Trans. on Sig. Proc. Supplement on Secure Media (2005)
2. Blech, M., Peter, C., Stahl, R., Voskamp, J., Urban, B.: Setting up a multimodal database for multi-study emotion research in HCI. In: Proceedings of the HCI International Conference, Las Vegas (2005)
3. BSI, 2004. British standards institute: BS 7986: Industrial process measurement and control – Data quality metrics (2004) Available from BSI Customer Services email: orders@bsi-global.com
4. Cowie, R., Douglas-Cowie, E., Tsapatsoulis, N., Votsis, G., Kollias, S., Fellenz, W., Taylor, J.G.: Emotion recognition in human computer interfaces. IEEE Signal Processing Magazine (January 2001)
5. Fasel, B., Luettin, J.: Automatic Facial Expression Analysis: A Survey. Pattern Recognition 36(1), 259–275 (2003)
6. Henry, M.P.: Self-Validating Sensors – Towards Standards and Products. Automatione e Strumentazione (2001)
7. Hudlicka, E.: Affect Sensing and Recognition: State-of-the-Art Overview. In: Proceedings of the 2005 HCI International Conference, Las Vegas. vol. 11 (2005) CD-ROM ISBN 0-8058-5807-5
8. Oertel, K., Schultz, R., Blech, M., Herbort, O., Voskamp, J., Urban, B.: EmoTetris for Recognition of Affective States. In: Proceedings of the 2005 HCI International Conference, Las Vegas (2005) CD-ROM. ISBN 0-8058-5807-5
9. Oudeyer, P.: The Production and Recognition of Emotions in Speech: Features and Algorithms, Sony CSL Paris (2003)
10. Paeschke, A.: Prosodische Analyse emotionaler Sprechweise, Logos Verlag, Berlin (2003)
11. Peter, C., Ebert, E., Beikirch, H.: A Wearable Multi-Sensor System for Mobile Acquisition of Emotion-Related Physiological Data. In: Proceedings of the 1st International Conference on Affective Computing and Intelligent Interaction, Beijing, 2005, pp. 691–698. Springer Verlag Berlin, Heidelberg, New York (2005)
12. Peter, C., Oertel, K., Kaiser, R., Schultz, R., Göcke, R., Voskamp, J., Urban, B.: The EREC sensor system for affect detection - application studies and results. Special session on emotion in HCI at the HCI International Conference, Beijing (2007)

13. Picard, R.W.: Affective Computing for HCI. In: Proceedings of the 8th International Conference on Human-Computer Interaction: Ergonomics and User Interfaces-Volume I, Lawrence Erlbaum Associates, Inc., Mahwah, NJ (1999)
14. Picard, R.W., Vyzas, E., Healey, J.: Toward Machine Emotional Intelligence - Analysis of Affective Physiological State. IEEE Trans. on Pattern Analysis and Machine Intelligence, vol. 23(10) (October 2001)
15. Tonti, M.: The influence of emotional and cognitive processes in the definition of speech rate. 37th annual meeting of the Society for Psychotherapy Research. Edinburgh (June 21–24, 2006)