H. Reiterer & O. Deussen (Hrsg.): Mensch & Computer 2012
München: Oldenbourg Verlag, 2012, S. 263-272

MediaBrain: Annotating Videos based on Brain-Computer Interaction

Alireza Sahami Shirazi, Markus Funk, Florian Pfleiderer, Hendrik Glück, Albrecht Schmidt

VIS, University of Stuttgart

Abstract

Adding notes to time segments on a video timeline makes it easier to search, find, and playback important segments of the video. Various approaches have been explored to annotate videos (semi) automatically to summarize videos. In this research we investigate the feasibility of implicitly annotating videos based on brain signals retrieved from a Brain-Computer Interface (BCI) headset. The signals provided by the BCI can reveal different information such as brain activities, facial expressions, or the level of users' excitement. This information correlates with scenes the users watch in a video. Thus, it can be used for annotating a video and automatically generating a summary. To achieve the goal, an annotation tool called MediaBrain is developed and a user study is conducted. The result reveals that it is possible to annotate a video and select a set of highlights based on the excitement information.

1 Introduction

This paper investigates the feasibility of implicitly annotating videos based on the information provided by the BCI (Brain Computer Interface). Adding annotations to time segments on a video timeline makes it easier to search, find, and playback important segments of the video. Various approaches have been explored to annotate videos (semi) automatically in order to summarize videos. Annotations are either defined automatically by analyzing and processing videos or explicitly by users/annotators.

The brain is the center of the nervous system. Brain signals can reveal different information that correlates with scenes the users watch in a video. The BCI enables communication between a brain and an external device. Several methods such as Magnetoencephalography (MEG), functional Magnetic Resonance Imaging (fMRI), and electroencephalograms (EEG) can acquire the brain signals. EEG is one of the most popular methods due to the ease of use. It uses non-invasive number of electrodes spread over the scalp to measure signals arising from neural activities. This allows to measure and analysis the brain neural activity without complex medical procedures. The brain neural activity varies based on the mental and cognitive activities. The BCI has been primarily used in the medical and clinical fields. However, with advantages in technologies low-cost commercial wireless EGG headsets are available for other purposes, e.g., games (Myeung-Sook et al., 2010).

In this project we utilize brain signals as implicit inputs for annotating video time segments and extracting a set of highlights. Generally, the user interaction with the computer is categorized into explicit and implicit interactions. In the explicit interaction the user aims for a certain action and expects certain results. While in the implicit interaction the user is not primarily aimed to interact with the computer, but the computer understands the interaction as an input (Schmidt, 2000). Researchers have used various resources such as the eye movements (Santella et al., 2006, Buscher et al., 2008) for implicit interactions. Here, we explore implicit video annotating based on the information provided by the BCI. Brain signals can reveal different information such as facial expressions or the level of excitement. This information can be used for annotating a video and generating a summary. To achieve the goal, we develop an annotation tool and conducted a user study. Based on the authors' knowledge, this is the first step toward using brain signals for annotating videos.

2 Related Work

There are various methods available to measure brain signals. The MEG (Magnetoencephalography) technique maps the brain activity by recording magnetic fields produced by electrical currents occurring naturally in the brain. fMRI (functional Magnetic Resonance Imaging) measures brain activity by detecting associated changes in the blood flow. The EEG (Electroencephalograms) technique measures brain voltage fluctuations resulting from ionic current flows within the neurons. It measures six signals each with different frequency ranges (alpha: 8–13 Hz, beta: 13–30 Hz, gamma: 30–100+ Hz, delta: up to 4 Hz, theta: 4–8 Hz, mu: 8–13 Hz). With the advantages in new technologies, commercial EEG headsets are recently available. This provides an opportunity to utilize these headsets in laboratory studies. The two most popular ones are NeuroSky[1] and Emotive EPOC[2] headsets. The NeuroSky devices have two electrodes and distinguish neutral and attentive mental states with 86% accuracy (NeuroSky, 2009). The Emotiv EPOC headset has 14 data collecting electrodes and 2 reference electrodes. It detects various facial expressions, level of engagement, frustration, mediation, and excitement.

Various research projects have used the both headset in order to use brain signals in different context. *ThinkContacts* is an application allows users to call a contact in an address book by using brain signals as inputs. It uses the NeuroSky MindSet to measure the degree of attention each contact gets in an address book to find out which contact to call (Perkusich et al.,

[1] http://neurosky.com/ (accessed March 2012)

[2] http://www.emotiv.com/ (accessed March 2012)

2011). *Neurowander* is also a BCI game using brainwaves as inputs for a game (Myeung-Sook et al., 2010). Mostow et al., 2011, used the EEG headset to collect and assess cognitive information from students while reading different texts. Crowley et al., 2010, assessed and reported the suitability of the NeuroSky MindSet to measure and categorize a user's level of attention and mediation. Furthermore, Petersen et al., 2011, demonstrated the ability to distinguish emotional responses reflected in scalp when viewing pleasant and unpleasant pictures. Yasui, 2009, proposed a technique for measuring the psychophysiological status of the human and associated applications based on brain signals. He analyzed the mental state of a car driver and showed that the pattern while driving was changed by a specific activity such as when talking on a mobile phone. We refer to Lotte et al., 2007, for an overview on classification algorithms for EEG-based brain-computer interfaces.

Researchers have also used the ERP (Event Related Potential) wave for interaction with a system. The ERP wave is the brain response that is directly the result of a thought or perception. The P300 wave is the famous ERP elicited in the process of decision-making. Kanoh et al. used the P300 signal for controlling the mouse course. It works by cycling through the eight possible directions around the current cursor position. When the signal is triggered, the mouse moves into the desired direction (Kanoh et al., 2011). Li et al. developed a P300-based keyboard that basically works by cycling through all letters until the desired one is reached (Li et al., 2009). *NeuroPhone* is a system that uses ERP signals obtained from the EPOC headset to select a contact from an address book on an iPhone and dial the number (Campbell et al., 2010). One of the brain signal monitoring usages in the medical domain is sleep monitoring to assist patients with sleeping disorders. Although the brain signals during sleep are quite weak and difficult to read (Garg et al., 2011).

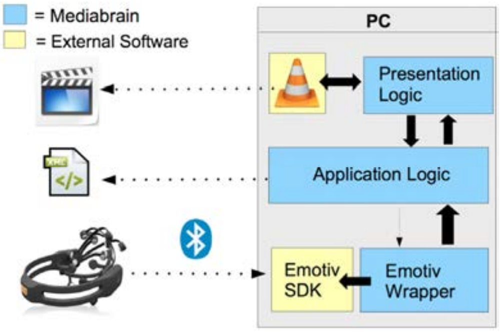

Figure 1:MediaBrain Architecture

Regarding annotating videos, researchers have investigated various automatic or semiautomatic approaches. Yamamoto et al., 2008, used the social activity, i.e., users' comments and weblogs for annotating videos. Nagao et al., 2002, provided an annotation tool allowed users to easily create annotations including voice transcripts, video scene descriptions, and visual/auditory object descriptions. Sahami Shirazi et al., 2011, used an iconic interface on the mobile phone for sharing opinions during sport events, annotating the events, and detecting highlights. Nakamura et al., 2008, explored affective response to understand video commenting systems. Various algorithms also tried to automatically annotate videos (Wang et al., 2009, Lavrenko et al., 2004). Saur et al. developed a tool, which automatically annotated basketball videos based on their content (Saur et al., 1997).

None of the previous work utilized emotional information for annotating a video. Emotional reactions of a video viewer to scenes in a video are correlated to what happens in a scene. In our research we used the brain signals acquired from the Emotiv EPOC headset to annotate a video and find highlights. In order to achieve this, we developed a prototype described in the next section.

3 MediaBrain

To annotate videos with emotional information acquired from the brain, the Emotiv EPOC headset was used and an annotation tool, called MediaBrain, was developed.

3.1 Annotation tool

We used the Emotiv EPOC headset to obtain the brain signals during watching a video. The headset uses the EEG technique and has 14 electrodes and 2 reference electrodes. It transmits data to the computer via a Bluetooth connection. The SDK (Software Development Kit) that comes with the headset provides following measurements: facial expressions, level of engagement, frustration, mediation, and excitement. The headset has also a built in gyroscope that detects the user's head orientation. The sampling rate is 128 samples/second.

A video annotation tool called MediaBrain is developed as an application for a personal computer. The tool includes the open source VLC[3] video player for fully controlling the video events such as playback, pause, or stop a movie. It is implemented in Visual C++. The MediaBrain tool consists of three layers (Figure 1). The first layer (Emotiv Wrapper) establishes a Bluetooth connection with the EEG headset and handles the user's brain signals. It uses the EPOC's SDK to retrieve the brain information. The second layer (Application Logic) records and stores the data in an XML file. It also tags the data with the video timestamp. The third layer (Presentation Logic) includes a wrapper around the VLC Media Player and controls the video. In the current version just excitement values are recorded and used to annotate a video. However, it is easily possible to extend the tool and use other parameters for the annotation. After gathering the information, the tool uses the XML file to identify, extract, and play scenes highlighted based on the excitement information.

[3] http://www.videolan.org/vlc/ (accessed March 2012)

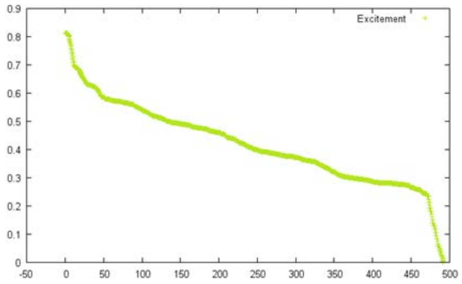

Figure 2:Sorted user excitement values

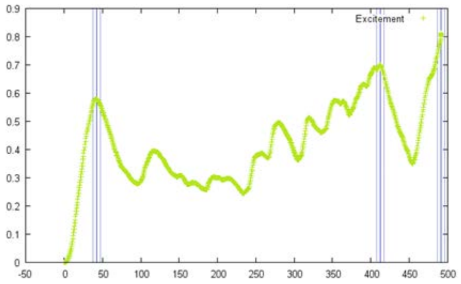

Figure 3:User excitement graph with calculated highlights

3.2 Annotation Algorithm

An algorithm is developed to extract the highlights based on the recorded information. The algorithm requires two parameters: length of a highlight (*L*) in seconds and maximum number of highlights (*N*). In the first step all highlights are sorted in descending order based on the excitement value (see Figure 2). Then, a highlight segment is detected. To do so, the scene that has the maximum excitement value together with the ±*L/2* seconds is extracted. The other excitement values in this time segment are excluded for further calculation. This procedure is continued till the maximum number of highlights (*N*) is calculated or no more data is available. If *N* is not provided, all available points are extracted. Figure 3 depicts the excitement graph with calculated highlights. The algorithm is described in Table 1.

```
N = maximum number of highlights;
L = length of a highlight;
excitement_array = sort the excitement data in a descending
                   order;
For 1 till N:
    item = Select first item in excitement_array;
    highlight_start_time = item.timestamp – L/2;
    highlight_end_time   = item.timestamp + L/2;
    in excitement_array remove data from highlight_start_time
    till highlight_end_time;
```

Table 1: The pseudo code describes the algorithm for annotating the scenes with the excitement values.

4 User Study

A user study was conducted to evaluate the MediaBrain tool and assess the feasibility of annotating the video based on the excitement information provided by the Emotiv EPOC headset.

4.1 Apparatus

We had 11 participants (7 male, 4 female, average age 23.2) for the user study. All participants were students recruited via mailing lists and university's forums. In the first step, each participant was asked to answer a questionnaire about the demographics. Then, we continued with watching a video. The participant wore the headset and started watching a short animation movie, called Big Buck Bunny[4]. We assured that all the 16 electrodes had a good connection to the scalp. We selected this movie as it had few funny scenes that should result in a distinct excitement graph. The movie's length was 10 minutes and shown on a 40" display. Figure 4 shows the setup of the user study. The participant wore the EPOC headset while watching the animated movie on the big screen. The experimenter was able to check the live data on the second monitor.

Along with the excitement data obtained implicitly from the EPOC headset, the users were asked to explicitly specify their excitement while watching the movie. Hence, we extended the MediaBrain tool in a way that users could state their excitement by pressing a button. During the study we asked the participants to press the button every time they believed they were excited about a scene. This information was stored together with the video timestamp in an XML file and used later for the evaluation. At the end of the study, the participants filled in another questionnaire and provided qualitative feedback about their experience during the study. The study took approximately 30 minutes for each participant.

```
N = maximum number of highlights;
L = length of a highlight;
excitement_array = detect points where the gradient changes;
For 1 till N:
    item = Select first item in excitement_array;
    highlight_start_time = item.timestamp – L/2;
    highlight_end_time   = item.timestamp + L/2;
    in excitement_array remove data from highlight_start_time
    till highlight_end_time;
```

Table 2: The pseudo code describes the updated algorithm for annotating the scenes with the excitement values.

4.2 Result

Based on the demographic questionnaire, 70% of the participants daily used their computer for watching videos. None of the participants took part in any user study related to the BCI or used any type of BCI headsets. Only one participant saw the animation shown in the study before.

[4] www.bigbuckbunny.org (accessed March 2012)

The results revealed that the participants pressed the button (specified their excitement explicitly) 13 times on average. There were six scenes where 85% of the users pressed the button. The rest of the explicit highlights were widely spread. The six scenes included mainly unexpected actions in the movie, led to a surprise and excitement.

We also investigated the correlation between the explicit and implicit excitement information from users. We took each explicit input from users and checked whether this input matched with a highlight detected by the algorithm. The results showed that with *L*=5 seconds only 27% of explicit inputs matched with the implicit excitements. With *L*=10 seconds the result was 36%. Further investigation revealed that the user inputs were on average 10 seconds earlier than the local maximum excitement values. Interestingly, the user inputs matched with the points where the excitement level started increasing (changes in gradient). However, we expected that the explicit inputs located on the local maximums (peaks) in the excitement graph (see Figure 5). Based on the Model Human Processor (Card, et al., 1986) the total cycle time of processors in the human's cognitive system, namely the *perceptual,* the *cognitive,* and the *motor processor* is approximately 300 msec. On the other hand, it might be delays the headset has. However, based on the headset manufacture documents no delay is reported. Therefore, we updated our algorithm in a way that the points where the excitement level started increasing were considered as highlights (see Table 2). Based on the updated algorithm we analyzed the data again. The results showed that with the new algorithm 65% of the users inputs overlapped with the highlights extracted via the algorithm.

The qualitative feedback showed that all users were relaxed during the study and enjoyed watching the movie. None of the users found the interaction with the EPOC headset inconvenient or disturbing. Also, all users mentioned that the explicit defining of excitements was not distracting them from concentrating on the movie. 77% of the users stated that they could imagine using the system in daily situations, like in front of the TV.

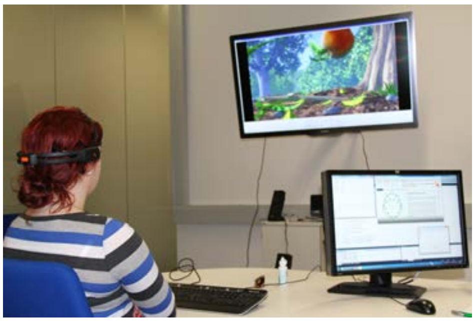

Figure 4: User taking part in the study

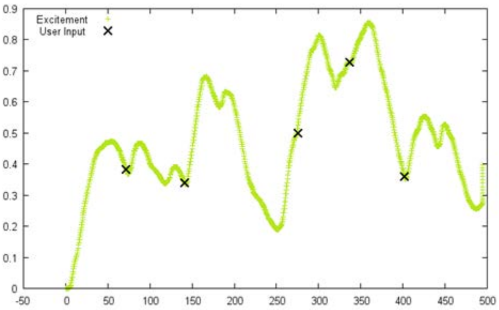

Figure 5:User-Highlights mapped to the excitement graph

5 Discussion & Conclusion

In this research we aimed at annotating videos based on excitement information acquired from a BCI headset. Hence, an annotation tool was developed and a user study was conducted.

The results reveal correlations between the scenes in the movie and the excitement level acquired from the BCI. Videos can be annotated with excitement information obtained from the EPOC headset and highlights can be extracted. Though the local maximums in the excitement graph correlates with the highlights in the video, but these are not the moments users believe they are excited. Moments which users think they get excited are the points where the excitement value starts increasing (gradient changes) in the excitement graph. Therefore, it is important to consider these points instead of the peaks in the excitement graph for annotating a video and generating a summary.

This research shows it is feasible to implicitly annotate a video based on excitement information and generate a set of highlights. Users emotional reactions during watching a video are rich resources for implicitly annotating and extracting important scenes in a video. Furthermore, the annotation can be used to automatically generate a summary of a video. Annotating a video with different emotional information gives us this opportunity to create various summaries based on different criteria.

We are currently planning to investigate whether other emotional information such as frustration or facial expressions can be used for annotating a video. Additionally, sharing the emotional reactions during watching a movie between non-collocated viewers might result in an increase in the connectedness and awareness among them.

Acknowledgment

This project is sponsored by the German Research Foundation (DFG) in the Cluster of Excellence in Simulation Technology (EXC 310/1) at the University of Stuttgart.

References

Buscher, G., Dengel, A., & van Elst, L. (2008), *Query expansion using gaze-based feedback on the subdocument level*. In Proceedings of ACM SIGIR'08 conference on Research and development in information retrieval , ACM, S. 387–394.

Campbell, A., Choudhury, T., Hu, S., Lu, H., Mukerjee, M.K., Rabbi, M. & Raizada, R.D.S. (2010). *Neurophone: brain-mobile phone interface using a wireless eeg headset*. In Proceedings of the second ACM SIGCOMM workshop on Networking, systems, and applications on mobile handhelds, S. 3-8

Card, S., Moran, T., and Newell, A (1986). The model human processor. In Kenneth R. Bo, Lloyd Kaufman, and James P. Thomas, editors, Handbook of Perception and Human Performance, chapter 45. John Wiley and Sons.

Crowley, K., Sliney, A., Pitt, I & Murphy, D. (2010). *Evaluating a brain-computer interface to categorise human emotional response.* In Advanced Learning Technologies (ICALT), 2010 IEEE 10th International Conference, 276-278

Garg, G., Singh, V., Gupta, J.R.P., Mittal A.P. & Chandra, S. (2011). *Computer assisted automatic sleep scoring system using relative wavelet energy based neuro fuzzy model.* In WSEAS Transactions on Biology and Biomedicine,

Kanoh, S., Miyamoto, K. & Yoshinobu, T. (2011). A p300-based bci system for controlling computer cursor movement. In Conference proceedings: Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Conference, volume 2011, S. 6405

Lavrenko, V., Feng, S.L. & Manmatha, R. (2004), *Statistical Models for automatic video annotation and retrieval*. In Proceedings. IEEE International Conference on Acoustics, Speech, and Signal Processing

Li, Y., Zhang, J., Su, I. Chen, W., Qi, Y., Zhang, J. & Zheng, X. (2009). *P300 Based BCI messenger.* In Complex Medical Engineering. CME. ICME International Conference, S. 1-5

Lotte F., Congedo M., Lecuyer A., Lamarche F. & Arnaldi B. (2007), *A review of classification algorithms for EEG-based brain–computer interfaces.* In Journal of Neural Engineering.

Mostow J., Chang K., Nelson J. (2011). *Toward exploiting eeg input in a reading tutor.* In Artificial Intelligence in Education, S. 230-237

Myeung-Sook, Y., Joonho, K., Sunghoon, K. (2010). *Neurowander: a bci game in the form of interactive fairy tale.* In Proceedings of the 12th ACM international conference adjunct papers on Ubiquitous computing, Ubicomp '10 Adjunct, S. 389-390

Nagao, K., Ohira, S. & Yoneoka, M (2002), Annotation-based multimedia summarization and translation. In Proceedings of the 19th international conference on Computational linguistics - S. 1-7

Nakamura, S., Shimizu, M. &Tanaka, K (2008). *Can social annotation support users in evaluating the trustworthiness of video clips?* In Proceedings of the second Workshop on Information credibility on the web, S. 59–62.

NeuroSky (2009), *NeuroSky's eSense™ meters and Detection of Mental State*. Neurosky, Inc.

Perkusich, M.B., Rached, T.S. & Perkusich, A. (2011). *Thinkcontacts: Use your mind to dial your phone.* In Consumer Electronics (ICCE), 2011 IEEE International Conference, S. 105-106

Petersen, M.K., Stahlhut, C., Stopczynski, A., Larsen, J.E. & Hansen, L.K. (2011*), Smartphones Get Emotional: Mind Reading Images and Reconstructing the Neural Sources.* In Proceedings of fourth International Conference on Affective Computing and Intelligent Interaction

Rebolledo-Mendez, G., Dunwell, I., Martínez-Míron, E. Vargas-Cerdán, M., De Freitas, S., Liarokapis, F. & García-Gaona, A. (2009). *Assessing Neuroskys usability to detect attention levels in an assessment exercise.* In Human-Computer Interaction, New Trends, 149-158

Sahami Shirazi, A., Rohs, M., Schleicher, R., Kratz, S., Müller, A. & Schmidt, A. (2011), *Real-time nonverbal opinion sharing through mobile phones during sports events.* In Proceedings of ACM Conference on Human Factors in Computing Systems, New York, NY, USA, S. 307-310.

Santella, A., Agrawala, M., DeCarlo, D., Salesin, D., & Cohen, M. (2006), *Gaze-based interaction for semi-automatic photo cropping.* In Proceedings of ACM Conference on Human Factors in Computing Systems ’06 , ACM, S. 771–780.

Saur, D.D., Tan,Y.-P. Kulkarni, S.R. & Ramadge, P. J. (1997), *Automated analysis and annotation of basketball video*. In Proceedings of SPIE's Electronic Imaging conference on Storage and Retrieval for Image and Video Databases V, S. 176-187

Schmidt, A. (2000), *Implicit human computer interaction through context.* In Personal and Ubiquitous Computing, S. 191-199

Wang, M., Hua, X.-S., Hong, R., Tang, J., Qi G.-J, Yan Song (2009), *Unified Video Annotation via Multigraph Learning.* In IEEE Transactions on Circuits and Systems for Video Technology, Volume 19, Issue 5, S.733 - 746

Yamamoto, D., Masuda, T., Ohira, S. & Nagao, K. (2008), *Video Scene Annotation Based on Web Social Activities*. In IEEE MultiMedia Volume 15, Number 3: S. 22-32

Yasui, Y. (2009), *A brainwave signal measurement and data processing technique for daily life applications.* In Journal of physiological anthropology, Volume 28, Number 3: S. 145-150

Contact Information

VIS, University of Stuttgart, Pfaffenwaldring 5a, 70569, Stuttgart Germany
{alireza.sahami, albrecht.schmidt}@vis.uni-stuttgart.de
{funkms, pfleidfn, glueckhk}@studi.informatik.uni-stuttgart.de