BCI for Physiological Text Annotation

Oswald Barral
Helsinki Institute for
Information Technology HIIT
Dept. of Computer Science,
University of Helsinki
oswald.barral@helsinki.fi

Ilkka Kosunen
Helsinki Institute for
Information Technology HIIT
Dept. of Computer Science,
University of Helsinki
ilkka.kosunen@helsinki.fi

Tuukka Ruotsalo
Helsinki Institute for
Information Technology HIIT
Dept. of Computer Science,
University of Helsinki
tuukka.ruotsalo@helsinki.fi

Michiel M. Spapé
Helsinki Institute for
Information Technology HIIT
Dept. of Computer Science,
Aalto University
sovspape@hiit.fi

Manuel J. A. Eugster
Helsinki Institute for
Information Technology HIIT
Dept. of Computer Science,
Aalto University
manuel.eugster@hiit.fi

Niklas Ravaja
Helsinki Institute for
Information Technology HIIT
Dept. of Social Research,
University of Helsinki
Dept. of Computer Science,
Aalto University
niklas.ravaja@aalto.fi

Samuel Kaski
Helsinki Institute for
Information Technology HIIT
Dept. of Computer Science,
Aalto University
Dept. of Computer Science,
University of Helsinki
samuel.kaski@hiit.fi

Giulio Jacucci
Helsinki Institute for
Information Technology HIIT
Dept. of Computer Science,
University of Helsinki
Dept. of Computer Science,
Aalto University
gulio.jacuccil@helsinki.fi

ABSTRACT
Automatic annotation of media content has become a critically important task for many digital services as the quantity of available online media content has grown exponentially. One approach is to annotate the content using the physiological responses of the media consumer. In the present paper, we reflect on three case studies that use brain signals for implicit text annotation to discuss the challenges faced when bringing passive brain-computer interfaces for physiological text annotation to the real world.

ACM Classification Keywords
H.5.3. Information Search and Retrieval: Relevance feedback; H.1.2. User/Machine Systems: Human factors

BCIforReal'17, March 13 2017, Limassol, Cyprus

ISBN 978-1-4503-4901-7/17/03...$15.00

DOI: http://dx.doi.org/10.1145/3038439.3038445

Author Keywords
Passive BCI, Physiological Text Annotation, Physiological Computing

INTRODUCTION
Annotating media content with descriptive metadata allows better content management and analysis. Content annotation has traditionally been a critical backbone of many digital media services, as it typically implies some explicit action from the users. The major benefits of content annotation are that, on the individual level, it allows for enhanced content delivery, namely through recommendation and personalization, taking into account the user preferences. [1, 8].

In previous studies, we have studied passive brain-computer interfaces (passive BCIs) for implicit annotation of text content. Passive BCIs refer to the specific branch of BCI that do not require the user to engage in any specific task to elicit specific brain patterns that can then be tracked but rather aims at capturing naturally-evoked reactive states. Our approach has mostly been to use BCI to enhance user-machine interaction, by using brain signals as additional implicit inputs to the system.

In this paper, we discuss the applicability of BCI in the real world, for physiological text annotation, by reflecting on three case studies. The first case study intended to study under controlled conditions the use of passive BCIs as a source for implicit text annotation, namely relevance annotation. Then, in the second case study, we applied the learning to close the loop, making use of the implicit relevance annotation to recommend potentially interesting information to the users. Here by relevance, we mean the semantic relevance of content to the search topic.

Finally, the third case study is an attempt to bring physiological text annotation to more naturalistic and less intrusive setups, studying affect annotation. Exploring the three case studies allows us to reflect on them, and raise a discussion around the benefits, inconveniences and major challenges faced when bringing passive BCIs outside of the laboratory.

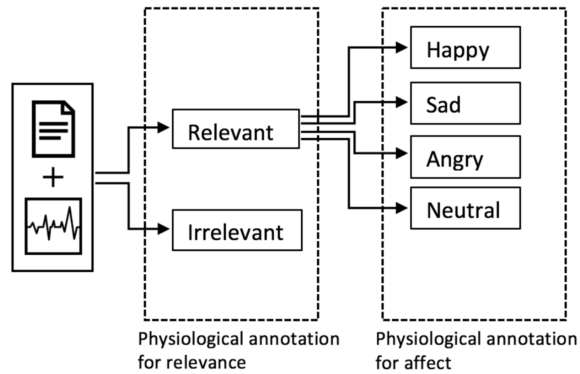

Figure 1. The physiological text annotation framework. ***Figure reproduced from*** **[6].**

PHYSIOLOGICAL TEXT ANNOTATION

The most common practice to annotate media content is through explicit interaction. While this has the advantage of being very accurate, it requires experts, or the users themselves to engage in the annotation. In the first case, having experts to annotate the content is usually costly and expensive. In the second case, it has been shown that users are not always willing to interrupt their task to provide feedback even when they are aware that this could lead to benefits in the subsequent interactions with the system [10].

Explicit feedback mechanisms may suffer from biases and do not allow for continuous monitoring. Other ways to generate content annotation is by implicitly monitoring the users, and using *implicit signals* to infer the annotations. For instance, behavioral signals have largely been explored as sources for implicit feedback, including scrolling patterns, clicks, and dwell time, among others [11, 14]. A clear example of the use of implicit sources for annotation to better target content to its users is YouTube, which bases the recommendations not only on the users' likes or subscriptions to channels (explicit signals), but also on the time users spend watching videos (implicit signal) [5].

Another very promising mechanism to implicitly annotate digital content is by using physiological signals [13, 6, 3, 16, 7]. Physiological recordings have the advantage that they can capture the emotional state of the user, which gives important clues to the users' preferences [15]. We recently introduced the concept of physiological text annotation, which refers to the practice of associating physiological measures to text content to infer characteristics of the user information needs and affective responses [4]. The physiological text annotation framework considers physiological recordings to annotate text content in two stages of the information consumption process. First, physiological recordings are used to annotate the relevance of the information items. Then, once the item is deemed as relevant, it is consumed, possibly eliciting several affective responses. In this second phase, physiological recordings are regarded for affect annotation (see figure 1).

A specific case that we have been studying is to use brain signals as sources for physiological text annotation. Next, we present three case studies to illustrate the potential of passive BCIs for physiological text annotation. However, bringing BCIs to the real world, where physiological text annotation will be most beneficial, is still an unresolved issue, and we, therefore, reflect on the case studies to analyze which are the most critical issues to consider when aiming to bring physiological text annotation through BCIs into the real world.

CASE STUDY 1: TERM-RELEVANCE PREDICTION FROM BRAIN SIGNALS

In [6] we investigated the first steps for relevance prediction through brain signals in a controlled laboratory experiment. Forty participants were presented with keywords and were asked to assess their relevance to a given topic, while their brain activity was measured through a 32-channel EEG headset.

In the experiment, the keywords were presented one at a time, and the users were asked to rate them relevant or irrelevant by pressing either the M-key or the X-key on a standard keyboard. The participants were instructed to give the relevance judgement as soon they made the decision whether the presented keyword was relevant or not. The next keyword was shown immediately after the feedback was given by the participant.

After pre-processing the signal to remove noise and artifacts, we extracted both frequency-based and ERP-based features from the EEG signals. The former were aimed at capturing the change in the signals across the whole time the words were presented to the user. The later aimed at capturing changes in the signals for a specific short time window when the participant made the relevance judgment.

The main purpose of the experiment was to study where and how binary relevance judgments of text affect neural activity, and if it is possible to predict such relevance judgments from it. We, therefore, extracted a set of different frequency- and ERP- based features to capture all the data that was potentially related to the relevance judgment. Features were time-locked to the stimulus onset and offset.

We then built several classification models, using different combinations of features sets, and studied which were the best features and combinations of features that better captured the changes in the EEG signals due to the relevance judgments. Results showed that the best performance was achieved using features capturing Alpha and Gamma activity, as well as ERPs. In this case, the mean classification accuracy for the participants was of 0.56, which represents an 11.72% improvement over the random baseline (random baseline was 0.5, as we designed the experiment in a balanced setup).

CASE STUDY 2: RECOMMENDING INFORMATION BY RELEVANCE INFERRED FROM BRAIN SIGNALS

In [7] we introduced a brain-relevance paradigm for information filtering. We used the knowledge gained from the case study 1 and formulated the hypothesis that relevance prediction of individual words from EEG signals can be utilized to automatically recommend a set of relevant documents to the user (see figure 2).

Fifteen participants read sentences from the English Wikipedia dataset. The experiment consisted of eight reading tasks. In each task, the participants were given two topics of which they choose on as relevant and another as the irrelevant. Each task was divided into six trials corresponding to the first six sentences extracted from the two topics. In each trial, the words of a sentence from the relevant topic was shown sequentially followed by words from a sentence in the irrelevant article. Thus, the words were shown one at a time while the brain signals of the users were monitored, and each word was classified as either relevant or irrelevant based on the brain signals. We then used real-time relevance-prediction of terms to query the Wikipedia dataset to extract further relevant documents for the user. Results showed that we were able to predict the relevance of words in context (words within a sentence), significantly outperforming the random baseline for 86.6% of the participants.

The final goal of the experiment was to use the relevance prediction for document retrieval and targeted recommendations. We queried the Wikipedia database with the words predicted as relevant, weighted with their tf-idf values (a standard measure in information retrieval that indicates the importance of a word within a corpus)[9]. Also, we queried the database using randomized predictions on the words. We then asked experts to evaluate the results in both cases. The results regarding document retrieval performance showed that the information gain for 83.3% of the participants was significantly greater when querying with the real predictions compared to the randomized feedback (considering only participants for which relevance prediction on the words level was significantly greater than random).

CASE STUDY 3: STUDYING AFFECTIVE ANNOTATION IN A NATURALISTIC INFORMATION CONSUMPTION SETUP

In an ongoing study, we aim at studying BCIs targeting the other end of the physiological annotation framework, which consists in affective annotation (see figure 1). The aim is also to study whether the affective annotation can be carried out in less controlled experimental setups.

In the experiment, twenty-four participants browsed comic strips from the 9GAG website. We then study how well we

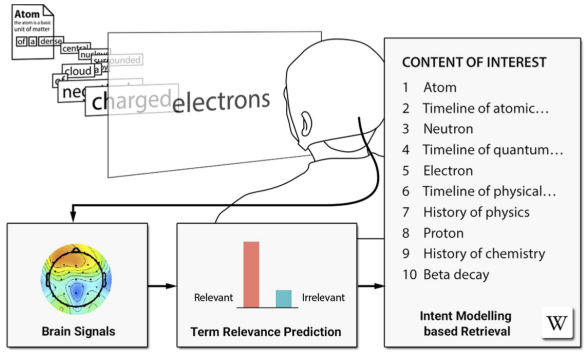

Figure 2. The brain-relevance paradigm for information filtering. ***Figure reproduced from*** **[7].**

can predict the humorousness of comic strips from single channel EEG recording, around the Pz location, which is as far as possible from frontal interference (eyes), and sits closest to the highest power of alpha activity [12]. Humor is a very subjective matter, and we do not use any predefined ground-truth for the comic; instead, each user gives explicit feedback after each comic to indicate whether they found the strip humorous or not.

We generated frequency-based features that capture the change in the signals across the whole time the participants read the comic strips. The analysis is still in process, but interestingly, preliminary results seem to indicate that simple frequency-based features (mostly features capturing Gamma activity) can annotate correctly up to 70% of the comic strips read by the users as funny or not funny. One noteworthy aspect of this study is that we did not need to use any blink or eye movement artifact removal in this experiment.

IMPLICATIONS FOR BCI OUT OF THE LABORATORY

Information retrieval, text annotation, and recommender systems are some of the promising fields for BCI to have a great impact when applied in a real world setting as it can be satisfactorily utilized even accounting for an error margin. For instance, a recommendation system that takes into account a relevance annotation generated through BCI -having a 60% probability that an item was relevant to the user- to improve the recommendation towards the more likely relevant item, will most likely still improve the overall search performance of the user. However, for instance, an explicit control system that would have 40% error rate would most likely feel completely unusable.

The -textual- information retrieval task is well suited for real-life BCI as the users usually consume information when they are stationary and often seated, thus minimizing the sources of movement artifacts. Similarly one could argue that text content, whether in the laboratory or real world, is less problematic for BCI than some more active content, such as movies, in that there is little unrelated visually-evoked neural activity that could degrade BCI performance. The drawback of textual content is that it might evoke weaker responses than some more exciting stimuli making not only the noise but also the signal weaker.

A major aspect to keep into account when bringing BCIs to the real world is the intrusiveness of the recordings. At this point in time, it is still hard to imagine end-users using laboratory grade EEG recording devices which are cumbersome and require expertise and time to set up (e.g. the EEG setups used in case studies 1 and 2). In the case study 3, on the other hand, the EEG was recorded from a single electrode which makes it much more feasible for real world deployment. The recording devices used in most laboratory experiments are prohibitively expensive as well often difficult for a layperson to use. Possible workarounds are to use affordable and easy-to-use devices that are available in the market (such as the Emotiv Epoc) which have been shown to provide reliable results at least in the detection of some event-related EEG responses [2]. However, at this stage, the quality of the acquired signal from these devices is still far from the high-end devices, and one should keep this into account when designing BCIs for the real world.

The case studies presented in this paper represent different types of BCI paradigms which include varying levels of challenge when applied to the real world settings. The milliseconds-level time-lock approaches (e.g.required for ERP analysis) used in case studies 1 and 2 require very precise synchronization between the recording device and the stimulus. Such precision in the recordings and synchronization is likely to be challenging in real-world BCI systems, as it needs very strict requirements for the hardware and software used to present the stimulus (textual content) as well as the networking between the stimulus device and the recording device (for example Bluetooth connection latencies can be difficult to quantify). A possible solution is to use frequency-based approaches (as in case study 3), which operate in the seconds- instead of milliseconds-level, and therefore is more robust to possible software and hardware synchronization irregularities. One could argue that the problems of synchronization are a greater issue for all annotation techniques that are based on the event-related responses, such as the classic P300, but less stringent on frequency-based approaches that could quantify the EEG activity. For example, the precision of the synchronization needed when quantifying the amount of alpha wave activity during the time a user spends reading an article, which could take several minutes, would be less stringent.

Eye movements are a source of artifacts and how to deal with them in real-world conditions is one of the main aspects to consider. In use case 1 and 2, we presented the words one at a time, in the center of the screen, minimizing eye movements, which allowed for cleaner data collection. However, under more realistic setups, it is unlikely that words can be presented one at a time in a fixed position, and therefore the possible artifacts introduced by eye movements need to be handled. One could use video-based eye tracking devices or even electrooculography (EOG, more invasive) to separately detect eye movements, in order to implement automatic rejection of EEG artifacts. An additional use of the measured eye movements could be to study ERPs associated to the user's fixations (i.e., fixation-related potentials), in which case the latency of the event-related potential with respect to the eye movements becomes an additional factor to be taken into account [17].

Another practical problem when deploying physiological text annotation paradigms through BCI into the real world is that the BCI system must have access, or be linked with the media content being consumed in some or the other way. Currently, the existing content does not provide any means of doing so. That is, when the user browses a web news site, the BCI system has no information on what is being displayed in the browser at any given time. A possible approach to solving this issue is to design an overlay proxy server, which was done in case study 3. In this way, instead of accessing external websites directly, users go through a proxy server that kept track of when the content changes in the browser, allowing synchronization with the EEG recording system. This allows the users to access any web content while still allowing the system to be aware of what content is being presented at any given time.

Finally, the last component required to implement physiological text annotation through BCI in the real world is the actual recommender system or search engine which is in charge of utilizing the physiological annotations for relevance and affect to better target the results to the users. The most obvious approach is to build a recommendation system that includes physiological text annotations for its recommendations, handling the uncertainty inherent in the BCI-based text annotations (as in case study 2). Another, more versatile but less accurate, approach is to send the annotations to an existing system, for instance, by masking it as manual input. For example, after the user has read a webcomic, the system would automatically press the funny or unfunny feedback button on the page as automatically generated from the BCI system, releasing the user from the tedious task of providing explicit feedback to the system.

CONCLUSION

This paper presented three case studies that use passive BCIs for physiological text annotation. We then reflected on them to identify the major challenges to face when bringing these paradigms outside the lab, in real-world applications.

ACKNOWLEDGMENTS

We thank Khalil Klouche for designing Figure 2. This work has been partly supported by the Academy of Finland (278090; 305739; Multivire, 255725; 268999; 292334; 294238; and the Finnish Centre of Excellence in Computational Inference Research COIN), and MindSee (FP7 ICT; Grant Agreement 611570).

REFERENCES

1. G. Adomavicius and A. Tuzhilin. 2005. Toward the next generation of recommender systems: a survey of the state-of-the-art and possible extensions. *Knowledge and Data Engineering, IEEE Transactions on* 17, 6 (June 2005), 734–749. DOI: http://dx.doi.org/10.1109/TKDE.2005.99

2. N. A. Badcock, P. Mousikou, Y. Mahajan, P. de Lissa, J. Thie, and G. McArthur. 2013. Validation of the Emotiv EPOC®EEG gaming system for measuring research quality auditory ERPs. *PeerJ* e38 (2013).
3. Oswald Barral, Manuel J.A. Eugster, Tuukka Ruotsalo, Michiel M. Spapé, Ilkka Kosunen, Niklas Ravaja, Samuel Kaski, and Giulio Jacucci. 2015. Exploring Peripheral Physiology As a Predictor of Perceived Relevance in Information Retrieval. In *Proceedings of the 20th International Conference on Intelligent User Interfaces (IUI '15)*. ACM, New York, NY, USA, 389–399. DOI: http://dx.doi.org/10.1145/2678025.2701389
4. Oswald Barral, Ilkka Kosunen, Tuukka Ruotsalo, Michiel M. Spapé, Manuel J. A. Eugster, Niklas Ravaja, Samuel Kaski, and Giulio Jacucci. 2016. Extracting relevance and affect information from physiological text annotation. *User Modeling and User-Adapted Interaction* 26, 5 (2016), 493–520. DOI: http://dx.doi.org/10.1007/s11257-016-9184-8
5. James Davidson, Benjamin Liebald, Junning Liu, Palash Nandy, Taylor Van Vleet, Ullas Gargi, Sujoy Gupta, Yu He, Mike Lambert, Blake Livingston, and Dasarathi Sampath. 2010. The YouTube Video Recommendation System. In *Proceedings of the Fourth ACM Conference on Recommender Systems (RecSys '10)*. ACM, New York, NY, USA, 293–296. DOI: http://dx.doi.org/10.1145/1864708.1864770
6. Manuel J.A. Eugster, Tuukka Ruotsalo, Michiel M. Spapé, Ilkka Kosunen, Oswald Barral, Niklas Ravaja, Giulio Jacucci, and Samuel Kaski. 2014. Predicting Term-relevance from Brain Signals. In *Proceedings of the 37th International ACM SIGIR Conference on Research & Development in Information Retrieval (SIGIR '14)*. ACM, New York, NY, USA, 425–434. DOI: http://dx.doi.org/10.1145/2600428.2609594
7. Manuel J. A. Eugster, Tuukka Ruotsalo, Michiel M. Spapé, Oswald Barral, Niklas Ravaja, Giulio Jacucci, and Samuel Kaski. 2016. Natural brain-information interfaces: Recommending information by relevance inferred from human brain signals. *Scientific Reports* 6 (December 2016), 38580. DOI: http://dx.doi.org/10.1038/srep38580
8. Jonathan L. Herlocker, Joseph A. Konstan, Loren G. Terveen, and John T. Riedl. 2004. Evaluating Collaborative Filtering Recommender Systems. *ACM Trans. Inf. Syst.* 22, 1 (Jan. 2004), 5–53. DOI: http://dx.doi.org/10.1145/963770.963772
9. Karen SpÃd'rck Jones. 1972. A statistical interpretation of term specificity and its application in retrieval. *Journal of Documentation* 28 (1972), 11–21.
10. Diane Kelly and Xin Fu. 2006. Elicitation of Term Relevance Feedback: An Investigation of Term Source and Context. In *Proceedings of the 29th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '06)*. ACM, New York, NY, USA, 453–460. DOI: http://dx.doi.org/10.1145/1148170.1148249
11. Diane Kelly and Jaime Teevan. 2003. Implicit Feedback for Inferring User Preference: A Bibliography. *SIGIR Forum* 37, 2 (Sept. 2003), 18–28. DOI: http://dx.doi.org/10.1145/959258.959260
12. Wolfgang Klimesch. 2012. Alpha-band oscillations, attention, and controlled access to stored information. *Trends in cognitive sciences* 16, 12 (December 2012), 606âĂŤ617. DOI: http://dx.doi.org/10.1016/j.tics.2012.10.007
13. Yashar Moshfeghi and Joemon M. Jose. 2013. An Effective Implicit Relevance Feedback Technique Using Affective, Physiological and Behavioural Features. In *Proceedings of the 36th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '13)*. ACM, New York, NY, USA, 133–142. DOI: http://dx.doi.org/10.1145/2484028.2484074
14. M. Soleymani and M. Pantic. 2012. Human-centered implicit tagging: Overview and perspectives. In *Systems, Man, and Cybernetics (SMC), 2012 IEEE International Conference on*. 3304–3309. DOI: http://dx.doi.org/10.1109/ICSMC.2012.6378301
15. Marko Tkalcic, A. Kosir, and Jurij Tasic. 2011. Affective recommender systems: the role of emotions in recommender systems. In *Proc. The RecSys 2011 Workshop on Human Decision Making in Recommender Systems*. Citeseer, 9–13.
16. Erin Treacy Solovey, Daniel Afergan, Evan M. Peck, Samuel W. Hincks, and Robert J.K. Jacob. 2015. Designing Implicit Interfaces for Physiological Computing: Guidelines and Lessons Learned Using fNIRS. *ACM Trans. Comput.-Hum. Interact.* 21, 6, Article 35 (Jan. 2015), 27 pages. DOI: http://dx.doi.org/10.1145/2687926
17. Markus A Wenzel, Jan-Eike Golenia, and Benjamin Blankertz. 2016. Classification of Eye Fixation Related Potentials for Variable Stimulus Saliency. *Frontiers in Neuroscience* 10 (2016), 23. DOI: http://dx.doi.org/10.3389/fnins.2016.00023