Adaptive games for cognitive training: Lessons measuring arousal with EEG

Daniel Sjölie
Umeå University
Umeå, SE-981 87, Sweden
daniel@cs.umu.se

CHI'13, April 27 – May 2, 2013, Paris, France.
ACM 978-1-4503-1952-2/13/04.

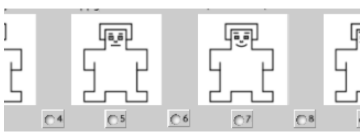
Abstract

Computerized cognitive training is an area where the use of computer games technology and methods has a great potential, for example, to address cognitive decline in an aging population. Adaptive games, in particular, are of great interest as the level of training has often been suggested as important for efficient training. An important part of any adaptive application is measuring and interpreting whatever the game should adapt to. In this paper we describe our work on using the Emotiv Epoc commercial EEG headset in order to measure and adapt to the user's level of arousal in two different applications. The first application is an adaptation of a classic cognitive training task (N-back) using game technologies to create a dynamic and (relatively) realistic version in a 3d-environment. The second application is a simple version of the classic space invaders game. In both applications EEG measurements recorded during initial training are used in a later phase to adapt the difficulty of the game automatically. While we managed to get this setup to work to a limited degree for some individuals, we failed to create a system where this method worked reliably across subjects and trials. In this paper, we describe what we tried, what worked, and some of the lessons we learned.

Ecological validity: A task is ecologically valid if it can be expected to give results that are valid in everyday life (our normal ecology). For example, training in a game world is ecologically valid if improvements in the game transfer to improvements in the real world.

Emotiv Epoc: Affordable, commercially available EEG headset for consumers.

SAM



The Self Assessment Manikin (SAM) may be used to present a visual scale of affective dimensions to users to get quick responses. We used scales for arousal and valence with 5 manikin images plus in-between scale steps.

Author Keywords
Adaptive games; brain-computer interfaces; cognitive training.

ACM Classification Keywords
H5.2 [Information Interfaces and Presentation]: User Interfaces, H.1.2 [Models and Principles]: User/Machine Systems — human factors; H.5.2 User Interfaces: Input devices and strategies; B.4.2 Input/Output Devices: Channels and controllers;

General Terms
Human factors, Measurement.

Introduction
Computerized cognitive training is an area with both great potential and significant challenges. Improvements on trained tasks are typical and transfer to similar task has been demonstrated (near transfer) [1] but transfer to general cognitive improvements (far transfer) is rare [3, 5]. In our work we focus on how the use of realistic and adaptive training can be used to improve the ecological validity and efficiency of computerized cognitive training. The use of adaptive games, in particular with realistic environments and 3d-graphics, is an interesting approach motivated by cognitive science theory.

Based on theoretical accounts of cognition and brain function in the context of cognitive training and realistic interaction [6, 7] the goal of the studies presented here was to adapt the training based on brain measurements. In particular, optimal training depends on an optimal level of prediction errors which should in turn be related to arousal. For example, according to activity theory any human activity is driven by a need to change something in the environment. That is, something important is not as we would prefer it to be, it does not match our "prediction" of the ideal world, and the mismatch arouses us to action.

In order to evaluate this approach we constructed a system where brain measurements were integrated into a game engine, and implemented two different tasks. Classifiers were trained to classify brain measurements based on reported arousal (measured using SAM, see column to the left).The goal was to use these classifiers to adapt the two tasks in real time based on brain measurements and the associated arousal and related level of prediction errors. The first task was directly inspired by established cognitive training and extended to be more realistic and adaptive in a 3d game environment. The second task instead focused on simplicity and game mechanics, within the same framework.

Methods
System components

The implementation was developed using the Panda3D game engine, and the Python programming language. A client server interface was integrated to receive data from the Emotiv Epoc EEG headset, together with extensive features for management and analysis of recorded data. The Enthough Python Distribution, building on scipy, numpy and matplotlib, was used extensively. Orange and scikit-learn were used for data classification.

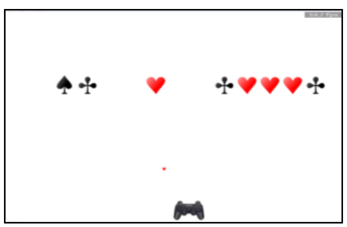
General task setup

The trials were divided into courses that each consisted of several one minute blocks. Changes and adaptations to the task were made between the one minute blocks.

Realistic N-back

This task was inspired by the dual N-back task used by Jaeggi et al. [3]. The stimuli to keep track of is **who** (which of the characters) is/was doing **what** (which animation) N steps back. This is a classic working memory task. Primary adaptations are changing N and the speed between motions.

Space invaders



This task is a simple version of the classic space invaders game. Primary adaptation is the speed at which aliens move (and descend).

Different courses used different initial settings as well as different methods for adaptation. For example, one course might start at N=1 and increase the difficulty automatically, to gather training data, while another course might start at N=2 and adapt according to classified brain measurements. See the fig. 1 for the setup of a course. Note that one course can contain an arbitrary number of blocks. The duration of states other than the *task* state were decided by the user, for example, by proceeding after reading instructions.

Figure 1. The structure of a course. SAM = Self Assessment Manikin.

All training examples that are used for classification are related to one such one minute block. See the column to the left for a short description of the two different tasks.

Classification

Classifiers were constructed and trained on examples from a number of blocks following a feature selection. We tried several methods of classification and feature selection during development and piloting. The method presented here is binary classification, high/low arousal, in combination with feature selection based on the information gain of discretized features. Orange was used to implement all of these steps. As a continuous output variable we used the probability of the "high arousal" class, giving a value between 0.0 and 1.0 corresponding to classified arousal.

Features

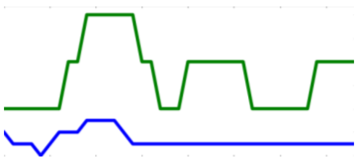
Features are based on the frequency spectra of the EEG-signal over certain time windows, as in previous work [2, 4]. We included the power in the standard frequency bands (delta, theta, alpha, beta and gamma), the frequency of the max peak for each band, and the power in the theta band relative to the power in the alpha and alpha+beta bands, respectively. We also included the phase correlation between the spectra for specific electrodes and the "expressive" signals output by the Emotiv library. For the frequency band power features we also generated features for a number of smaller frequency bands, such as 7Hz-8Hz, and 40Hz-44Hz. In total we had 1261 features.

In order to facilitate the continuous generation of these feature in a real-time application we used short time windows, 3.0 and 9.0 seconds, regenerated at shorter intervals to overlap. Using short overlapping time windows is a technique that has been used many times before. Lee and Tan [4] used even shorter overlapping windows (2 s, updated each second) but Grimes et al. [2] have demonstrated that performance of classification improves with increasing time windows. For the actual classification we used training examples built on descriptive statistics (primarily the median) of these short time features over the task part of a *block* (see figure 1).

Results

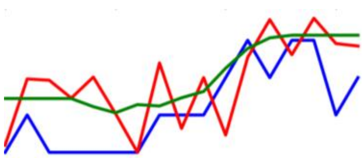
With the N-back task several users did not report the variation in arousal that was necessary to get good training data. This was not a problem with the space invaders task, reflecting that the explicit gaming aspects of the second tasks were more effective at eliciting varied arousal and engagement. Still, it should

Constant arousal



Reported arousal (blue) stays constant during second half although N (green) varies in an N-back task trial.

Classified arousal



Example of classified arousal (red) and reported arousal (blue) for a trial with correlation r=0.6. Notice the trend when classified arousal is smoothed (green).

Acknowledgements

Thanks to Swedish Brainpower, Umeå University and the Kempe foundations, JCK-2613, for funding.

Thanks to Lars-Erik Janlert and Johan Eriksson for contributions to the study and comments on the paper.

be noted that reported arousal did vary with N in most cases, to some degree. A Kruskal-Wallis non-parametric analysis of variance test show that reported arousal is different between N's with $p<0.05$ in 10 out of 13 N-back trials.

Since the attained classification performance was not good enough to support stable adaptation, as was our goal, the data collection was aborted before completed as planned. The gathered data covers 18 trials over 4 subjects: 13 trials with N-back and 5 trials with space invaders. Here we present results from this data in the form of the correlation coefficient (r) between reported arousal and classified arousal. For each trial we divided the data into two equal parts and used the first half as training data and the second half as test data. The correlation coefficients in question are for the test data compared to classifications based on the training data for the same trial.

In total there are 5 trials (28%) with $r>=0.6$. This can be broken down into N-back with 23% of 13 trials at $r>=0.6$ and space invaders with 40% of 5 trials at $r>=0.6$. These statistics are admittedly with low power but they may reflect the effect of increased arousal variation and engagement in the space invaders task.

Concluding remarks

The results presented here are preliminary, but they indicate that, while there are many challenges for adaptive games using commercial EEG headsets, it is possible given optimal conditions. They also provide support for further use of game methods and technologies to motivate and engage users. The disappointing performance of the intended adaptation may necessitate further research on basic tasks, but in the long run an optimal cognitive training application should combine strong points from both the tasks implemented here.

While we believe that much remains to be investigated using the system we have developed, the resources available to us for conducting such investigations in the near future are limited. We are interested in any collaboration taking advantage of our framework or the gathered data.

References

[1] Dahlin, E. et al. 2008. Transfer of Learning After Updating Training Mediated by the Striatum. *Science*. 320, 5882 (Jun. 2008), 1510–1512.
[2] Grimes, D. et al. 2008. Feasibility and pragmatics of classifying working memory load with an electroencephalograph. *Proceeding of the twenty-sixth annual SIGCHI conference on Human factors in computing systems*. (2008), 835–844.
[3] Jaeggi, S.M. et al. 2008. Improving fluid intelligence with training on working memory. *Proceedings of the National Academy of Sciences*. 105, 19 (May. 2008), 6829–6833.
[4] Lee, J.C. and Tan, D.S. 2006. Using a low-cost electroencephalograph for task classification in HCI research. *Proceedings of the 19th annual ACM symposium on User interface software and technology*. (2006), 81–90.
[5] Owen, A.M. et al. 2010. Putting brain training to the test. *Nature*. 465, 7299 (Jun. 2010), 775–778.
[6] Sjölie, D. 2012. Presence and general principles of brain function. *Interacting with Computers*. (2012).
[7] Sjölie, D. 2011. *Reality-based brain-computer interaction*. Umeå University, Department of Computing Science.

Participating Author Biography

Daniel Sjölie will attend the workshop. He has a licentiate degree in computing science with a focus on human-computer interaction and cognitive neuro-science in a context of interactive 3d-graphics.

He has several years of experience of virtual reality and computer graphics as a research engineer, and virtual reality and computer game technologies continue to be central themes in his (soon to be completed) PhD research. He has lectured and coordinated courses about game development and game design. Among his primary interests are the foundations in cognitive neuroscience for phenomena related to virtual reality and gaming, such as presence and engagement, and the application of such understanding to benefit, for example, cognitive training applications and serious gaming.