Combining EEG with Pupillometry to Improve Cognitive Workload Detection

Article *in* Computer · October 2015

DOI: 10.1109/MC.2015.314

2 authors:

David Rozado
Otago Polytechnic

Andreas Duenser
The Commonwealth Scientific and Industrial Research Organisation

Combining EEG with Pupillometry to Improve Cognitive Workload Detection

David Rozado, Otago Polytechnic

Andreas Dünser, CSIRO

Combining signals from traditional electroencephalography with those of novel physiological measures such as pupil dilation could enable more accurate and robust real-time monitoring of cognitive workload.

Using physiological data as real-time system input allows for the creation of new human–computer interaction (HCI) paradigms. By continuously monitoring the user's state, this approach can dynamically respond to the user and facilitate faster, smoother HCI. Traditionally, different neuroimaging modalities have been used to build functional brain–computer interfaces (BCIs), including electroencephalography (EEG), functional magnetic resonance imaging (fMRI), magnetoencephalography (MEG), functional near-infrared spectroscopy (fNIRS), electrocorticography, and intracortical recordings. Some of these are capable of detecting cognitive workload,[1,2] but the accuracy, robustness, and latency of these techniques are suboptimal and, in certain contexts, markedly inferior to self-reporting.[3]

Developing a robust and accurate BCI that can quickly and efficiently track cognitive workload remains a challenging task. We believe that traditional EEG-based cognitive workload detection can be improved by combining it with an additional information source like pupil dilation. To test this hypothesis, we combined pupil diameter with a set of EEG-derived features to improve the detection of cognitive effort. The pupil-diameter time series is captured using a remote video-based eye tracker. We found that using such multimodal techniques can improve the performance of physiological computing systems. Our work builds upon previous findings regarding pupil dilation and cognitive workload in situations that demand high levels of cognitive engagement.[4]

 0018-9162/15/$31.00

EEG AND COGNITIVE WORKLOAD

When exerting considerable mental effort, we tend to undergo delays in our information-processing capabilities. In extreme cases, we might become unresponsive to incoming sensory streams because the amount of input information exceeds the available mental resources to process the information. When experiencing excessively low mental effort, on the other hand, we can become bored with the task at hand, which, paradoxically, can increase the possibility of making mistakes. Thus, our working output is effectively a function of ongoing cognitive workload.

Monitoring a person's cognitive workload can be valuable for enhancing computing systems' operational safety, usability, and adaptive automation. Mechanisms that measure cognitive workload have been suggested at the behavioral, subjective, and physiological levels. However, behavioral and subjective measurements only provide indirect insights about cognitive workload, whereas physiological measurements are more practical and unbiased, providing direct measurements over time.

Several physiological measures have been used to detect variations in cognitive workload. Brain-related measures using neuro-imaging techniques such as fMRI or EEG directly monitor brain activity to infer cognitive workload. EEG is particularly useful because of its high temporal resolution, low cost, and portability. Studies have found correlations between cognitive workload and increases or decreases of theta (4–8 Hz) and alpha rhythms (8–12 Hz), depending on the user.[1] An early study found that theta activity increased while alpha and beta activity decreased with increased task difficulty.[2] The study's authors argued that theta activity in the left frontal parts of the brain might be correlated with general mental processing.

In our work, we used common spatial patterns (CSPs) to spatially filter raw EEG signals into derived time series that maximize cognitive state differences.[5] Because cognitive workload is characterized by variations in the power spectral density of different EEG bands, we applied a CSP algorithm to discriminate between two classifications of EEG recordings: data captured while users are carrying out mental arithmetic, and data captured while users are not engaging in any specific activity. The features extracted from the CSP procedure were then fed to a machine-learning classifier that finds the optimal separation between the cognitive state classes.

Several cognitive-state classification approaches based on EEG signals are discussed in the literature, such as auto-regressive neuronal networks, support vector machines (SVMs), or hidden Markov models.[6] Among these, simple linear discriminant analysis (LDA), which we used in our work, is successful for real-time BCI systems—it has very low computational requirements, is simple to use, and has high accuracy rates.[6]

COMBINING PUPILLOMETRY WITH EEG

In eye-related measurements of mental effort, the study of changes in pupil diameter has attracted considerable attention. The term *pupillometry* refers to monitoring pupil diameter using a camera pointed at the eye. Small changes in pupil diameter are linked to cognitive processes coupled with the activation of the brainstem (locus coeruleus).[4] Thus, measuring pupil-diameter oscillations can provide information about users' cognitive processes and engagement, such as their cognitive workload when performing mental-arithmetic tasks.[7]

However, changes in pupil diameter due to cognitive processes (ranging in size from 0.1 to 0.5 mm) can be obscured by changes in ambient illumination, stimulus brightness, or chemical substances that affect the pupil (ranging in size from 0.1 to 8 mm). This forces researchers to control ambient light and stimulus brightness during experiments involving pupil diameter.

We believe that combining changes in pupil diameter with EEG-derived features can improve the classification accuracy of a BCI designed to assess cognitive workload, compared to the same type of BCI that uses only EEG-derived features. For research by others on this subject, see the sidebar.

In a recent study, we found that the accuracy of an EEG-based motor-imagery BCI, which is often used in physical rehabilitation, can be improved by using pupillometry as an additional signal source.[8] The combined classifier achieved a movement-detection accuracy of 83.25 percent compared to 66.95 percent for EEG alone. In this article, we explore whether such an approach would work equally well for improving cognitive-workload classification.

EXPERIMENT: COGNITIVE-WORKLOAD ASSESSMENT

In this work, we designed an experiment to discriminate between two mental states: a cognitive-workload condition and a no-task control condition in which participants were asked to remain relaxed, doing nothing in particular. We recruited 23 participants—10 female and 13 male—between 15 and 48 years old (mean = 32, standard deviation =

SIMILAR RESEARCH

Previous studies in this field have yielded similar results to our studies. Kilseop Ryu and Rohae Myung used a combination of physiological measurements to evaluate cognitive workload in a dual-task scenario.[1] They combined electroencephalography (EEG), blink rate via electrooculogram (EOG), and heart-rate variability via electrocardiogram (ECG), and found that EEG performed better for inferring arithmetic-related workload, and EOG and ECG were better indicators for visual object tracking. A measure developed by combining these signals, however, correlated with difficulty variations in both kinds of tasks.

Thomas Fritz and his colleagues examined a combination of psychophysiological measures to predict task difficulty in software development.[2] They found that a classifier based on a combination of eye tracking, electrodermal activity, and EEG could predict task difficulty with 64.99 percent accuracy, whereas eye-tracking data alone achieved 69.16 percent accuracy.

Eija Haapalainen and his colleagues also compared combinations of various physiological measures, including EEG and pupillometry, to assess cognitive workload.[3] A combination of EKG and heat flux (rate of heat transfer) provided the best classification results, with 81.1 percent accuracy.

References

1. K. Ryu and R. Myung, "Evaluation of Mental Workload with a Combined Measure Based on Physiological Indices during a Dual Task of Tracking and Mental Arithmetic," *Int'l J. Industrial Ergonomics*, vol. 35, no. 11, 2005, pp. 991–1009.
2. T. Fritz et al., "Using Psycho-physiological Measures to Assess Task Difficulty in Software Development," *Proc. 36th Int'l Conf. Software Eng.* (ICSE 14), 2014, pp. 402–413.
3. E. Haapalainen et al., "Psycho-physiological Measures for Assessing Cognitive Workload," *Proc. 12th ACM Int'l Conf. Ubiquitous Computing* (UbiComp 10), 2010, pp. 301–310.

7.45). All but two participants were right handed and all had normal or corrected-to-normal vision.

For the cognitive-workload condition, we instructed participants to perform mental arithmetic, namely adding two random two-digit numbers without writing anything down or using a calculator. We asked participants to continuously carry out the arithmetic operation until a stop signal was given. The experiment was divided into three blocks of 5 minutes and 50 seconds each. During the first block, participants could familiarize themselves with the experimental setup. The remaining two blocks were used for actual model training and classification.

To control stimulus brightness, we told participants to fixate on a cross centered on a computer monitor for the length of each time block. To ensure that participants remained fixated on the cross, a beeping sound was triggered if they looked away. The room was illuminated with standard fluorescent light, and outside light was blocked so that ambient light remained stable for the experiment's duration.

Each experimental block consisted of 25 trials in a random sequence. Each trial started with an auditory cue (a synthetic voice either reading the arithmetic operation—for instance, 27 + 36 (for the cognitive-workload condition)—or the word "nothing" (for the no-task condition). Eight seconds after the cue, an auditory stop signal was given. Six-second breaks were provided after each trial so participants could relax and their brain electrical patterns could return to baseline levels. We instructed participants to minimize body movement to minimize potential artifacts in their EEG signals.

Software

We used the open source software packages EEGLab for data analysis and visualization and BCILAB for cognitive-workload detection,[9] and we used customized scripts to insert the pupil-diameter feature in the EEG-derived feature vector. We used the lab streaming layer (LSL) software library for networking streaming data, data collation, and time synchronization for EEG; gaze and pupil size time-series streams; and data streams containing stimulus events. The Simulation and Neuroscience Application Platform (SNAP) displayed the visual and auditory stimuli to the participants.

Hardware and data preprocessing

We used a Tobii X2-30 eye tracker to monitor gaze and pupil diameter. Gaze was sampled at 30 frames per second, and estimation accuracy was around 0.5 degrees of the visual angle. Blinks were filtered out of the raw gaze signal by rejecting gaze samples below a quality threshold. No filters were applied to the pupil-diameter signal.

We recorded EEG signals with a Biosemi ActiveTwo EEG amplifier with 32 Ag/AgCl active electrodes placed on the scalp using the 10-20 system. We recorded EEG signals at 512 Hz and then resampled them prior to data processing to 128 Hz. We used the common average reference (CAR) as a suitable EEG reference system for data analysis. This method uses the mean EEG signal of all electrodes as a reference baseline against which to compare the signal of each electrode.

Prior to calculating participant-specific CSP spatial filters, we filtered the re-referenced EEG signal in the 3–30 Hz band with a zero-phase forward/backward finite impulse-response filter. This frequency band contains the theta, alpha, and beta frequency bands. Spectral density oscillations in these four bands are associated with cognitive load.

Data analysis: common spatial patterns

The set of electrodes that permit detecting the maximum discriminatory power between experimental conditions varies between subjects due to differences in the anatomical properties of cortical folding and in the propagation of electrical signals in the brain. The CSP algorithm is suitable for constructing subject-specific spatial filters optimized for discriminating between two conditions in EEG data.

Figure 1 shows a set of CSP scalp projections generated for one participant in the user study. The CSPs are optimized for discriminating the cognitive workload condition from the no-task control condition.

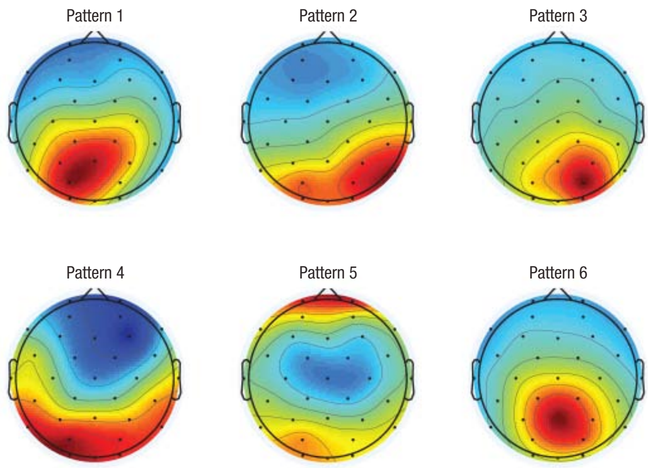

FIGURE 1. Common spatial pattern (CSP) maps. A set of CSP filters of a single participant in our study. The CSPs are optimized for discrimination of the cognitive workload condition from the no-task control condition.

Classification of Cognitive Workload

The raw EEG signal was processed by the CSP filters to generate the features used for cognitive-workload classification. For each condition (cognitive workload triggered by mental arithmetic, and no task), the variance of only a reduced number *m* of the signals was suitable for discrimination and used to construct the classifier. In our case, the CSP feature vector had a length of six, as three pairs of CSPs were computed for each experimental block. The signals that maximized the variance of the two experimental conditions were the signals associated with the largest eigenvalues λ_{nothing} and $\lambda_{\text{cognitiveLoad}}$ of the composite spatial covariance of the EEG time-series matrix. The derived feature vectors of the cognitive-workload and no-task conditions were used to estimate an LDA classifier.

We tested various descriptive features of the pupil-diameter samples for classification purposes: mean, maximum, minimum, standard deviation, skewness, kurtosis, slope, and power at the 0.1–0.2 frequency band. We combined all features with two different classification techniques: LDA and SVMs. The average pupil diameter (averaged over both eyes) was the best-performing feature in terms of classification accuracy. We used this feature in combination with EEG-derived features to test the performance of our hybrid approach. We derived the pupil-diameter feature by acquiring a trial baseline (–2,000 ms until the start cue) for each trial and then calculating the average pupil diameter from 500 ms after the start cue to 8,000 ms. We inserted average pupil diameter as an additional dimension in the CSP feature vector prior to LDA classification.

Figure 2 illustrates the different feature combinations used for cognitive-workload classification. All methods consisted of two stages: signal processing, which filtered the raw signals, and classification, which used different combinations of features. Figure 2a shows the classification of cognitive workload using pupil diameter only. We continuously updated a moving average of the pupil diameter and compared it to a reference baseline period right before the start of each trial. Figure 2b shows the method used to classify cognitive workload using EEG-derived signals and the CSP algorithm. Figure 2c illustrates our

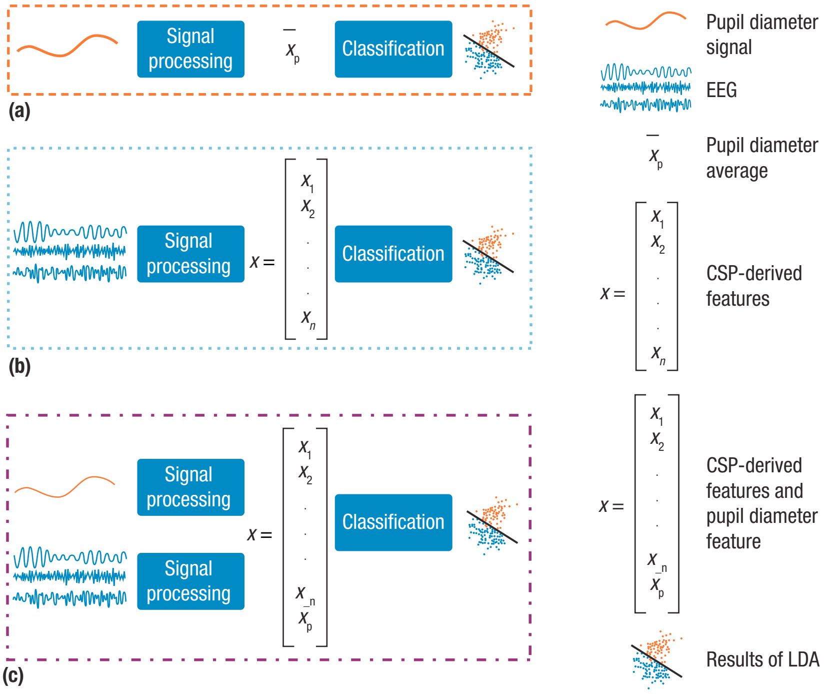

FIGURE 2. Feature combinations used for cognitive-workload classification. (a) Classification of cognitive workload using pupillometry only. (b) Classification of cognitive workload using just EEG. (c) Our method, which combines pupillometry with EEG. LDA stands for linear discriminant analysis, which is the classification algorithm used.

proposed method, which adds average pupil diameter to the EEG-derived feature vector before classification.

We tested two data-sampling approaches. In the first, we evaluated each block of experimental data using 10-fold cross-validation within the block, and reported the average error rate of both blocks. In the second, we used a block of data in its entirety for training and the remaining block for testing to derive the average error rate. The training set and test set assignments were then inverted, and we reported the average error rates on both test sets.

We analyzed the performance of the classification methods being tested using the error rate of the classifier, the number of bits transmitted per trial, the probability of guessing the correct classification due to chance, and the kappa coefficient.

Experimental results

Table 1 shows the results of the different classification methods using mean pupil diameter with or without EEG, as well as different classifier training approaches on the metrics used to evaluate performance. The kappa statistic compares the average error rate with the expected error rate (random chance). The information transfer rate is a simplified computational model based on the noisy-channel coding theorem.

We compared the classification error rate for detection of mental arithmetic–induced cognitive workload when using just the pupil-diameter feature, just the EEG features, and the combination of the pupil-diameter feature with the EEG features using repeated measures analysis of variance (ANOVA).

With the training and test block approach, we found a significant difference between the classification methods:

$$F(1.45, 31.89) = 9.39, p < .01, eta_p^2 = .30.$$

Degrees of freedom were adjusted with the Greenhouse–Geisser correction because of sphericity violation. Post-hoc comparisons using the Bonferroni correction showed that the error rate was significantly lower with the combination of EEG and pupillometry than with either EEG or pupillometry alone. The classification error rate did not differ significantly between EEG and pupillometry.

We also found a significant difference between classification methods with the 10-fold cross-validation training approach:

$$F(1.29, 28.29) = 5.12, p = .02, eta_p^2 = .19.$$

Post-hoc tests also showed that the error rate was significantly lower with the combination of EEG and pupillometry than with either EEG or pupillometry alone, and it did not differ significantly between EEG and pupillometry.

Figure 3 shows the average pupil diameter of all trials in a single block for the 23 participants in the user study. The blue line represents the no-task condition and the red line the cognitive-workload condition. Note that the pupil diameter variable was normalized using the time period of 2,000 ms prior to the trial start cue (the first 60 gaze samples in Figure 3) as a baseline. The figure shows that increased cognitive workload, triggered by the mental arithmetic, leads to enlargement of the pupil diameter (compared to the no-task condition), although the effect differs in size among participants.

Figure 4 illustrates the EEG signal (event-related spectral perturbation; ERSP) for all participants during the cognitive-workload condition in electrode Pz. The ERSP refers to changes in the power spectra of the EEG signal locked to a certain event (the start cue within each trial). In this manner,

TABLE 1. Experimental results.

Technique	Average classification error rate (%)	Standard deviation	Kappa coefficient	Information transfer rate (bits/trial)	Information transfer rate (bits/min.)
Pupillometry	26.1	12.2	0.48	0.17	1.28
Training/test blocks: EEG	33.2	13.7	0.34	0.08	0.63
Training/test blocks: EEG + pupillometry	18.1	10.6	0.64	0.32	2.38
Cross-validation within block: EEG	24.1	10.5	0.52	0.20	1.52
Cross-validation within block: EEG + pupillometry	17.0	8.5	0.66	0.34	2.56

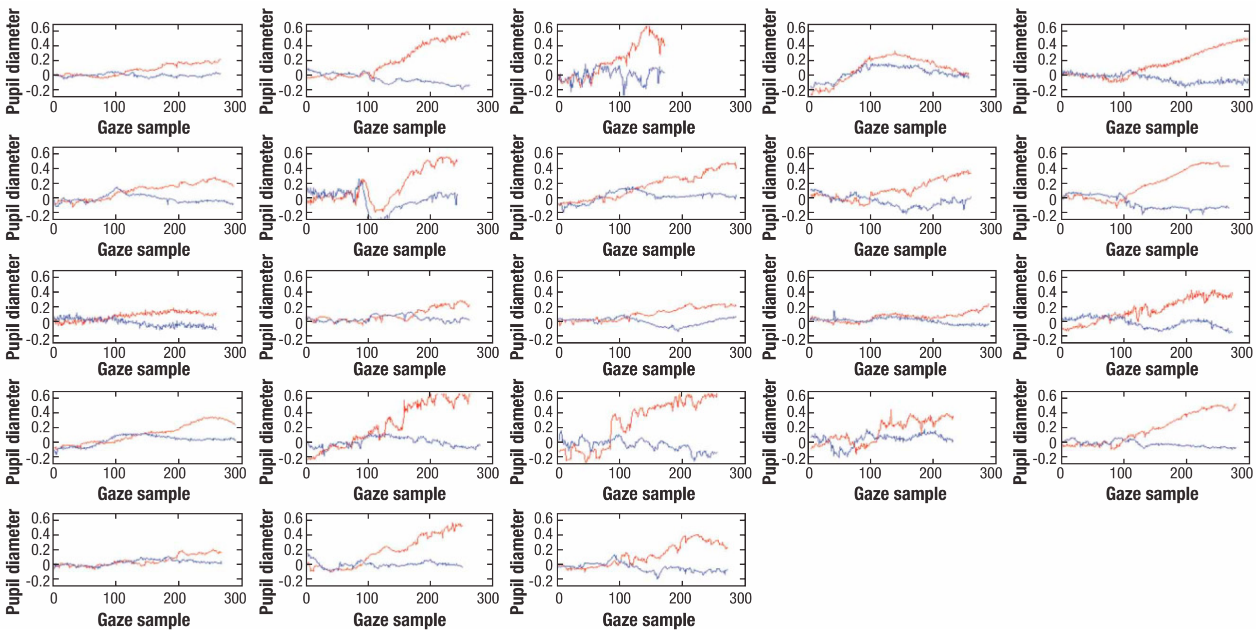

FIGURE 3. Pupil-diameter time-course averages. The pupil diameter is larger in cognitive-workload trials involving mental arithmetic (red line) than in the no-task trials (blue line), although the effect differs in size among participants. Pupil diameter is measured in mm.

spectral differences between conditions (in this case, the cognitive-workload and no-task conditions) can be compared in the frequency domain.

Note that the effect described in the literature of decreased power in the alpha band for the cognitive workload triggered by mental arithmetic[1] is not obvious for all participants in Figure 4. This is to be expected because of differences in cortical folds and the corresponding effects of volume conduction on electrical neural-signal propagation. Thus, event-related desynchronization corresponding to mental arithmetic might not be present at electrode Pz but could still be apparent elsewhere on the scalp. A consistent pattern for the theta band is even less obvious for the same reasons.

ADVANTAGES, LIMITATIONS, AND FUTURE RESEARCH

Our results show that combining a pupil-diameter feature with EEG-derived features can predict cognitive workload—elicited through simple arithmetic operations—with an error rate of 17 percent, compared to 26.1 percent for pupillometry alone and 24.1 percent for EEG alone. These results align with those from similar studies (see the sidebar), including our own previous study.[8] Although our new study might not be directly comparable to the others, which either used different setups or physiological measures, they all point to the benefits of combining several data sources to create better and

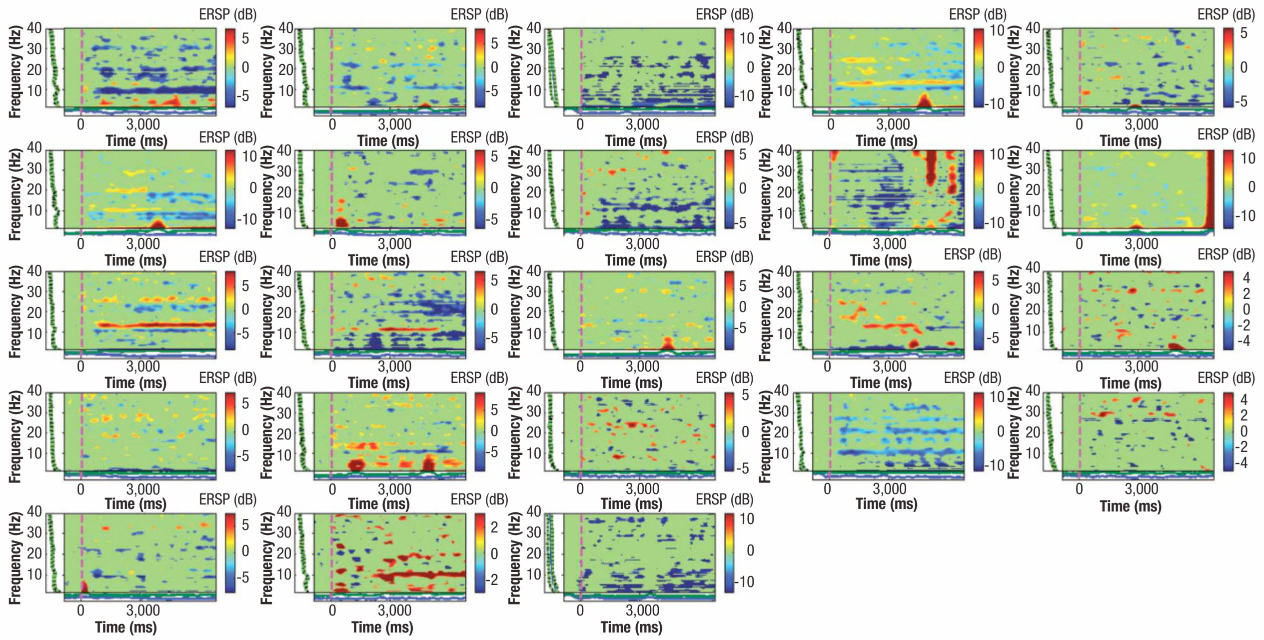

FIGURE 4. Frequency spectrum of EEG activity (event-related spectral perturbation; ERSP) for all participants during the cognitive-workload condition in electrode Pz. Event-related desychronization and synchronization in the alpha and theta bands are apparent in some but not all participants. dB is decibels.

more accurate physiological monitoring systems.

One issue when using physiological data is that it can be sensitive not only to internal cognitive processes but also, and often even more so, to external stimuli. To limit potential confounds due to ambient light changes, we used pupil dilation just prior to the start of each trial as a baseline. Thus, slow-wave frequency changes in ambient light can be accommodated and should not overly diminish the validity of the pupil diameter as a data source for physiological monitoring.

The brightness of the stimulus being gazed at might be more problematic. To keep this constant, we asked participants to fixate on a point throughout our experiment and used sound cues instead of visual cues to indicate the trial type (cognitive workload or no task). Although it is fairly standard practice in BCI research to limit eye-movement artifacts, we acknowledge that stimulus brightness can be problematic for pupillometry measurement. Future work could investigate ways to control for stimulus brightness, for example, by monitoring users' gaze on the screen and evaluating the screen brightness around the focal point. In this way, pupil diameter changes due to variations in stimulus brightness could be accounted for.

To enhance the generalizability of our results and demonstrate the potential of our proposed multimodal approach, we gathered EEG data under rather loose conditions. Our experimental blocks were relatively short—under six minutes each, well below most EEG or BCI data-gathering sessions reported in the literature, which often exceed 15 minutes. However, because our approach improves classification accuracy, we can afford a reduction in the amount of training data needed for calibration while providing high accuracy.

In experimental laboratory environments, researchers often manually curate data. These conditions, however, are very different from practical scenarios. In the real world, the performance of a real-time BCI cannot depend on expert data curation. However, eliminating artifacts in the data is common practice in EEG data analysis to optimize data quality for signal processing and classification algorithms. In our experiment, all raw trial data was used for training and test stages, regardless of oculomotor noise or the presence of EMG activity or other artifacts in the EEG data. Our proposed multimodal approach has great potential for real-time BCI systems because it achieves relatively high accuracy under suboptimal data-quality conditions.

BCI illiteracy is a well-described phenomenon[6] that refers to the inability of a certain percentage of the population to efficiently use a particular BCI paradigm (such as P300, mu rhythms, or slow cortical potentials). We chose participants randomly and did not filter out anyone, even if the EEG-based estimation of cognitive workload was poor. Thus, our results might be negatively biased by the inclusion of several subjects regardless of BCI illiteracy. However, adding pupillometry to the measurement arsenal can be advantageous for detecting cognitive workload in BCI-illiterate users.

The results obtained for cognitive-workload estimation using just the pupil-diameter feature are remarkably similar to classification performance using EEG-derived features. Although the difference between the combination of EEG and pupillometry and pupillometry alone was significant in both training approaches, the difference in error rate between pupillometry alone and EEG alone was not significant. This is very interesting because the pupil feature contains just 1 degree of freedom as opposed to the EEG signal, which accounts for 31 degrees of freedom. Furthermore, monitoring pupil size alone is much less cumbersome than monitoring EEG scalp channels.

An additional advantage of the pupillometry approach is that it does not require any training data because it uses a fixed threshold between classifications and a continuously updated normalization approach (the baseline period prior to the trial start cue). In addition, the pupil-diameter feature does not need manual data cleansing. But because pupillometry has its own specific limitations, the pupil-diameter feature and the EEG-derived feature complement each other in our proposed multimodal system.

At this stage, it is reasonable to assume that pupillometry has very limited granularity to distinguish between more than two states (such as high or low cognitive workload) on a single trial basis. EEG approaches, however, could more easily be adapted to distinguish between a gradient of cognitive-workload levels. We used a rather simple task for cognitive workload: adding two random two-digit numbers together. Although this task could induce workload levels so they are clearly distinguishable in the pupillometry (see Figure 3) and EEG features (see Figure 4), other, more diverse tasks should be tested in future studies. In addition, longer training and test blocks might have improved the EEG-based classification performance.

Considerable progress has recently been made in the detection accuracy of cognitive workload using EEG-based BCIs. However, performance remains suboptimal for the technology's use in real-world environments. Multimodal approaches that combine several electrophysiological signals, such as those from pupillometry and EEG, could help achieve greater reliability in detecting cognitive states. This would lead the way to better and more robust systems for direct, real-time measurement of cognitive workload, supporting better HCI and achieving greater user satisfaction.

ABOUT THE AUTHORS

DAVID ROZADO is a lecturer of information technology at Otago Polytechnic. His research interests include assistive technologies, brain–computer interfaces, machine learning, and human–machine interaction. Rozado received a PhD in computer science from Universidad Autónoma de Madrid. Contact him at david.rozado@op.ac.nz.

ANDREAS DÜNSER is a senior research scientist in the Cognitive Informatics Team at CSIRO (Commonwealth Scientific and Industrial Research Organisation). His research interests include the convergence of psychology and emerging interactive technology, human–computer interaction, research methods, and the design of new technologies for healthcare applications. Dünser received a PhD in psychology from the University of Vienna. Contact him at andreas.duenser@csiro.au.

REFERENCES

1. D. Grimes et al., "Feasibility and Pragmatics of Classifying Working Memory Load with an Electroencephalograph," *Proc. SIGCHI Conf. Human Factors in Computing Systems* (CHI 08), 2008, pp. 835–844.
2. A. Gundel and G.F. Wilson, "Topographical Changes in the Ongoing EEG Related to the Difficulty of Mental Tasks," *Brain Topography*, vol. 5, no. 1, 1992, pp. 17–25.
3. G.F. Wilson, "An Analysis of Mental Workload in Pilots during Flight Using Multiple Psychophysiological Measures," *Int'l J. Aviation Psychology*, vol. 12, no. 1, 2002, pp. 3–18.
4. B. Laeng, S. Sirois, and G. Gredebäck, "Pupillometry: A Window to the Preconscious?," *Perspectives on Psychological Science*, vol. 7, no. 1, 2012, pp. 18–27.
5. H. Ramoser, J. Muller-Gerking, and G. Pfurtscheller, "Optimal Spatial Filtering of Single Trial EEG during Imagined Hand Movement," *IEEE Trans. Rehabilitation Eng.*, vol. 8, no. 4, 2000, pp. 441–446.
6. F. Lotte et al., "A Review of Classification Algorithms for EEG-Based Brain–Computer Interfaces," *J. Neural Eng.*, vol. 4, no. 2, 2007, pp. R1–R13.
7. M. Pedrotti et al., "Automatic Stress Classification with Pupil Diameter Analysis," *Int'l J. Human-Computer Interaction*, vol. 30, no. 3, 2014, pp. 220–236.
8. D. Rozado, A. Duenser, and B. Howell, "Improving the Performance of an EEG-Based Motor Imagery Brain Computer Interface Using Task Evoked Changes in Pupil Diameter," *PloS ONE*, vol. 10, no. 7, 2015; doi: 10.1371/journal.pone.0121262.
9. C.A. Kothe and S. Makeig, "BCILAB: A Platform for Brain-Computer Interface Development," *J. Neural Eng.*, vol. 10, no. 5, 2013; doi: 10.1088/1741-2560/10/5/056014.

Computer Previous Page | Contents | Zoom in | Zoom out | Front Cover | Search Issue | Next Page Qmags THE WORLD'S NEWSSTAND®