

A statistics based approach to the use of EEG signals to measure workload

By Andrew Belyavin, Christopher Ryder, Blair Dickson

QinetiQ Ltd, Farnborough, UK

Abstract

Many military contexts impose high cognitive workload on the personnel involved in making decisions and this can cause critical degradation of performance. In 2002 the US Defense Advanced Research Projects Agency initiated the Augmented Cognition programme to develop technologies that could be employed in managing operator workload through direct measurement of cognitive activity and adapting system behaviour to improve performance. This paper describes the development of a method for calibrating the electroencephalogram to independent estimates of verbal and spatial workload so that cognitive load can be estimated directly from physiological signals over short time intervals in real time. A method for detecting and correcting artefacts using a kurtosis based Independent Component Analysis procedure is described and an approach for constructing an estimate of the sampling distribution is outlined. It is shown that the calibration procedure provides plausible results for constructing separate estimates of both verbal and spatial workload using established computer based tasks and this can be used to measure human workload in real time.

1. Introduction

Many military contexts impose high cognitive workload on military operators and this can cause poor decisions to be made under acute time pressure, undermining military effectiveness. In 2002 the US Defense Advanced Research Projects Agency (DARPA) established the Augmented Cognition programme to stimulate the development of technologies that can be employed in systems that can adapt to reduce operator workload when it is excessive. It is argued that relying on fully automatic systems can lead to brittleness at the edge of the design envelope of the system and this can be mitigated by maintaining the human operator(s) in the loop. The key elements of an Augmented Cognition system are the ability to directly measure the cognitive workload of the operator(s) and to instigate changes to system behaviour that reduce the load on the crew at these times. A number of approaches to workload reduction have been considered including adaptive automation, altering the flow of tasks requiring operator attention so that those of high priority are attended to first, and simple modification of the human computer interface so that high priority tasks can be addressed more rapidly. This approach could be applied in a range of military contexts such as the management of uninhabited vehicles in all environments, naval operations rooms and single seat aircraft for which mission-critical periods are characterised by very high cognitive workload.

Under the DARPA programme between 2002 and 2007 the QinetiQ team developed a method of analysing EEG signals and relating them to verbal and spatial workload while performing complex tasks as described in Belyavin *et al* (2007). The aim of the development was to provide a method that could be employed in Augmented Cognition systems to measure cognitive workload in real time and provide the criteria for changing system behaviour. The approach is based on a simple conceptual model of cognitive load, embodied in the Prediction of Operator Performance (POP) workload model (Belyavin and Farmer 2006). The key assumption underlying the POP model is that cognitive workload is directly determined by the amount of cognitive activity to be undertaken in a fixed time period. It also assumed that a

small number of cognitive activities can be undertaken in parallel, even if a considerable range of motor actions can be conducted simultaneously – most of us can walk and chew gum at the same time. Analysis of dual task experiments indicates that two important cognitive activities that can be conducted without significant interference are verbal and spatial tasks (Nicholls *et al* 2003). Using the assumptions of the POP workload model a method was developed based on detecting verbal or spatial load through direct measurement of brain activity using the Electro-encephalogram (EEG) and relating it to cognitive workload.

It is assumed that any specific brain activity measured from a fixed pattern of electrodes varies between different individuals and it is necessary to calibrate an individual's verbal or spatial activity pattern before the system can be used to measure cognitive load. To achieve individual calibration, a participant is tested with combinations of two tasks at various defined levels of verbal and spatial cognitive load and their EEG is recorded at the same time. It is assumed that the amount of time spent undertaking a verbal or spatial cognitive activity is directly related to cognitive load and based on this assumption the critical components of the EEG that relate to verbal and spatial load can be isolated.

The EEG signal as measured from the scalp is vulnerable to interference from extraneous signals such as the mains or muscle activity and steps must be taken to mitigate the impact of such interference. A key component of the process is the method used to clean the EEG in real time prior to detailed processing both at the calibration stage and at the live measurement stage. The aim of this paper is to describe the statistical basis both of the artefact removal procedure and the subsequent calibration procedure and to provide demonstration of their application to workload measurement.

2. Artefact removal – sampling distribution derivation

The aim of the artefact removal procedure is to identify anomalous signals in the EEG and remove them from subsequent processing. Over the last 20 years a number of powerful methods for detecting clusters in multivariate data have been developed under the general heading of Independent Component Analysis (ICA). The two main ICA methods either use sample entropy or sample moments higher than the second. It was decided in the current study to use a method based on sample kurtosis – the standardised fourth sample moment.

For a multivariate sample $\{\mathbf{X}\}$ with n cases and m variates the sample covariance matrix, \mathbf{S} is defined as

$$S_{ik} = \sum_{j=1}^n (X_{ij} - \bar{X}_i)(X_{jk} - \bar{X}_k) \quad (2.1)$$

where \bar{X}_i is the sample mean for the i^{th} variate and j labels cases. A matrix square root of \mathbf{S} can be defined, \mathbf{T} , such that $\mathbf{T}^T\mathbf{T}=\mathbf{S}$. The standardised sample with identity covariance matrix is then constructed as $\mathbf{Z}=\mathbf{T}^{-1}(\mathbf{X}-\bar{\mathbf{X}})$. The standardised kurtosis tensor is finally defined as

$$k_{ijkl} = \sum_{c=1}^n Z_{ic} Z_{jc} Z_{kc} Z_{lc} \quad (2.2)$$

A number of possible sets of useful directions can be defined from the fourth order tensor (see for example Moulines and Cardoso 1992) and a simple but effective variant is the contraction of the fourth order with the identity matrix

$$C_{ij} = k_{ijkl} \delta_{kl} \quad (2.3)$$

where δ_{kl} is the Kronecker delta and summation is implied over repeated suffixes. \mathbf{C} is a symmetric positive definite matrix, and for a Gaussian population \mathbf{C} is a diagonal matrix with elements $(m+2)$. Large eigenvalues point to clustered populations which include probable

outliers. Small eigenvalues point to uniform populations or small numbers of relatively equal sized clusters. The sample matrix C can be calculated rapidly even for large samples and the eigenvalue problem can be solved readily with standard methods, providing a fast ICA method. If it is to be used effectively for the removal of artefacts from signals it is necessary to estimate the sampling distribution of at least the largest eigenvalue of C for a Gaussian population.

It is not possible to construct an analytic form for the sampling distribution of C for a Gaussian sample of arbitrary size and dimension. Over the last 15 years there has been considerable progress in deriving asymptotic results for the sampling distribution of the largest eigenvalue of random matrices based on the identification of the limiting Tracy-Widom distribution (Tracy & Widom 1994); Johnstone (2001) demonstrated that the asymptotic sampling distribution of the covariance matrix with large dimension and sample size follows the Tracy-Widom distribution of order 1 for Gaussian populations. A two stage method was used to construct an estimate of the sampling distribution of the largest eigenvalue of C . First an asymptotic estimate of the distribution for large samples was constructed using Monte-Carlo simulation based on the asymptotic distribution of the elements of C . Using the asymptotic form as a limit, the sampling distribution for a range of sample sizes and dimensions was derived using Monte Carlo simulation of the construction of C for a range of sample sizes and dimensions. An interpolation formula in the combined set of simulation results was constructed using the following results.

THEOREM 1. *The asymptotic distribution of the largest eigenvalue of C for a Gaussian population for fixed m/n as $n \rightarrow \infty$ follows the Tracy-Widom law of order 1.*

Proof. Using standard results for the sampling moments of moments, the following statements can be made. The asymptotic distribution of all the elements of C is Gaussian. The asymptotic expected value of the diagonal elements of C is $(m+2)$ and the off-diagonal elements is 0. The asymptotic variance of the diagonal elements is $4(m+5)/n$ and the asymptotic variance of the off diagonal elements is $2(m+4)/n$. The asymptotic covariance between all the pairs of diagonal elements is $4/n$ and that between all other pairs of elements is 0. Based on these findings the asymptotic joint density of the elements of $D = \sqrt{n}(C - (m+2)I)$ where I is the identity matrix is given by

$$f(D) = \exp \left[-\frac{1}{2} \left\{ \frac{\text{Trace}(D^2)}{4(m+4)} - \frac{(\text{Trace}(D))^2}{8(m+2)(m+4)} \right\} \right] / F \quad (2.4)$$

where F is a normalising constant. This density can be transformed using standard theory to yield the asymptotic joint density of the scaled eigenvalues $\{\lambda_i\}$ as

$$f(\Lambda) = \exp \left[-\frac{1}{2} \left\{ \frac{\sum \lambda_i^2}{4(m+4)} - \frac{(\sum \lambda_i)^2}{8(m+2)(m+4)} \right\} \right] \prod_{i < j} (\lambda_i - \lambda_j) / G \quad (2.5)$$

where G is a normalising constant. This density has the same form, up to a term that is $O(m^{-1})$, as that for the covariance matrix. Since this term can be neglected as $m \rightarrow \infty$, the asymptotic distribution of the largest eigenvalue of C follows the Tracy-Widom law of order 1 with appropriate scale and location modification.

LEMMA 1. *The asymptotic distribution of the largest eigenvalue of D for a Gaussian population for $m=1$, $n \rightarrow \infty$ follow a Gaussian distribution with mean 0 and variance 24.*

This follows directly from standard results on the sampling moments of moments.

LEMMA 2. *The asymptotic expected value of the largest eigenvalue of D for a Gaussian population as $m, n \rightarrow \infty$ is $\sqrt{8m(m+4)}$*

Johnstone (2001) demonstrates that the asymptotic mean of the largest eigenvalue of a real Wishart matrix is $(\sqrt{(n-1)} + \sqrt{(m-1)})^2$. The expected value of the diagonal elements of a real Wishart matrix is n and that of the off-diagonal elements is 0; similarly the variance of the diagonal elements is $2n$ and that of the off diagonal elements is n . The variance of the component of D that is equivalent to the Wishart matrix is $4(m+4)$. Correcting for the scaling of the two matrices and subtracting the mean n from the expression given by Johnstone, the expected value is given as $\sqrt{8m(m+4)}$.

The results of Theorem 1 and Lemmas 1 and 2 were used to construct an interpolation formula for the limiting distribution of the largest eigenvalue of D. It was then assumed that the deviation of the sampling distribution from the limiting form could be expanded in powers of $\sqrt{1/n}$ for variations in sample size for fixed m , and in positive and negative powers of \sqrt{m} as m varies. Using a combination of these assumptions and the limiting form, an overall interpolation formula was constructed that provided reliable estimates of the upper percentage points of the largest eigenvalues of D for the region $150 > m > 0$, $n > m + 7$, $\sqrt{(m-1)/(n-1)} < 0.97$. Some example upper 0.5% points are displayed in Table 1.

Sample size	m=2	m=10	m=32	m=100
20	7.93	17.02	Undefined	Undefined
50	7.25	20.13	43.96	Undefined
125	6.08	17.80	47.76	116.82
1000	4.62	13.63	38.22	114.56

TABLE 1. Example upper 0.5% points for the largest eigenvalue of the contracted kurtosis tensor

3. Artefact removal – implementation

In the initial applications, the EEG data was gathered from an array of 14 electrodes at a frequency of 1024 kHz and the signal was analysed in epochs between 2 and 4 seconds long depending on the application, thus each block contained between 2048 and 4096 vectors. The first stage of the artefact removal process for a single block was to calculate the contracted kurtosis tensor using the outputs of all the EEG electrodes and to identify the directions for which the eigenvalues exceeded the upper 0.5% point for the largest root calculated using the interpolation formula outlined in Section 2. No more than 3 directions were processed. For each direction a formal model comprising a mixture of two Gaussian populations with a common mean of 0, but differing standard deviations was fitted to the standardised observations by maximum likelihood. The Bayes likelihood of each observation belonging to the population with larger standard deviation was calculated. Those observations for which the likelihood exceeded 0.5 were projected orthogonal to the corresponding eigenvector. This complete procedure was repeated for a maximum of 12 cycles for each block. If the observations in the block could not be corrected using this procedure so that the largest eigenvalue of the contracted did not exceed the upper 0.5% point, the process was terminated and the block was flagged as unreliable.

A sample of a single channel of EEG signal that is heavily contaminated with artefact (the RAW signal) and the result of the cleaning procedure (the CLN signal) is displayed in Figure

1. In this example 8 sweeps of the artefact removal procedure were conducted and considerable modification of the signal took place – a total of 11058 corrections in 4096 cases. As can be seen, the output of the cleaning is generally good, although there may be some increase in cleaned signal variance in the contaminated period between 0.7 and 1.5 seconds.

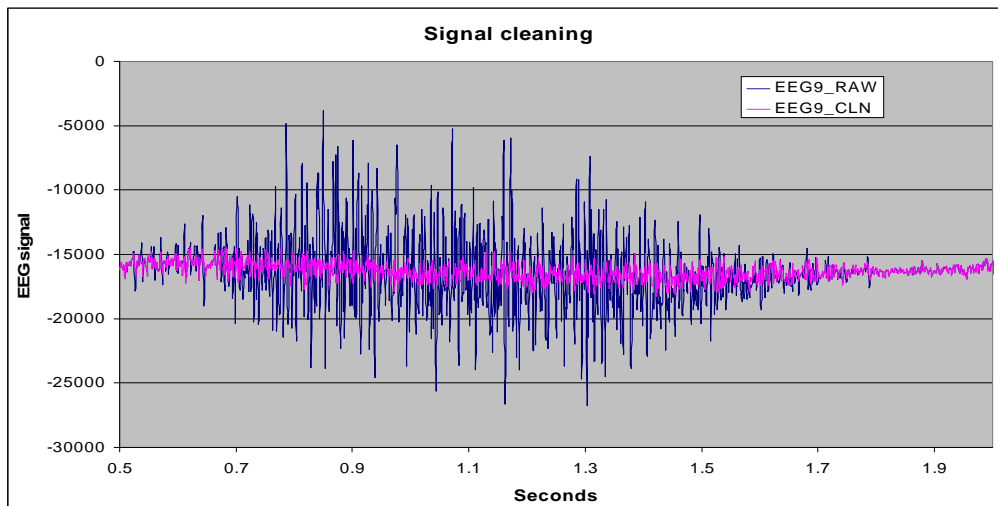


FIGURE 1. Plot of raw contaminated signal and cleaned signal that was passed as acceptable

4. Calibration of EEG parameters to workload

To relate the EEG to workload two laboratory tasks with known workload characteristics were used to calibrate the EEG response for each individual. There was substantial previous experience with two tasks that had well defined characteristics for which workload could be controlled: a two-dimensional compensatory tracking task and the Bakan vigilance task (Bakan 1959). Previous work indicated that compensatory tracking was dominated by spatial workload and that the Bakan task was dominated by verbal workload.

The calibration procedure comprised 10 tests each of 150 seconds' duration: 3 levels of the single tracking task; three levels of the single Bakan task; 4 levels of simultaneous tracking and Bakan task. More detail about the precise workload levels are described in Belyavin *et al* (2007). The principle underlying the manipulation of workload in the procedure is that if the time provided to the operator to make decisions is reduced, the cognitive workload will be increased according to a simple relationship derived from the POP model.

The workload rating assigned to a test is assumed to be uniform for the 150 second period. To meet this assumption, the recorded signals were analyzed in blocks of 4 seconds' length for tests involving a single tracking task and blocks of length corresponding to the nearest multiple of the inter-stimulus interval for tests involving a Bakan task (3.75 or 4.0 seconds). After each block had been subjected to artefact removal, the spectrum was calculated in each of nine frequency bands between 0 Hz and 100 Hz listed in Table 2. The cross spectra between the 14 scalp electrodes in the nine bands were derived to yield estimates of the coherence between electrode pairs and the gains of one electrode relative to another, providing 1638 inter-correlated variables describing the EEG in each block. Pairwise coherence was

selected as it should be insensitive to scale and location shifts in individual electrodes; relative gain should be insensitive to overall shifts.

Band	Range (Hz)
Delta	0.0 – 3.5
Theta	3.5 – 8.0
Alpha1	8.0 – 10.2
Alpha2	10.2 – 14.1
Beta1	14.1 – 20.0
Beta2	20.0 – 30.0
Gamma Low	30.0 – 47.0
Gamma Mid	53.0 – 70.0
Gamma High	70.0 – 100.0

TABLE 2. Spectral bands used in the calibration procedure

The best discriminants between different verbal and spatial workload values were then estimated using stepwise multiple regression of the workload values on the EEG parameters. First a stepwise up procedure was adopted where the criterion for including a variable was an F-value greater than 9.0. This was followed by a stepwise down procedure eliminating variables with an F-value less than 9.0. A summary of the results of applying this procedure to 8 subjects in a designed experiment involving either flying a simulated mission in a laboratory flight simulator or undertaking a simple logic task is provided in Belyavin *et al* (2007). It was found that verbal load was increased according to the EEG measures in the logic task and spatial workload was increased in the flying task.

Subsequent trials of the procedure indicate that it can be expected that R^2 values for good discriminants of verbal and spatial load should exceed 0.8. If this criterion is attained cognitive workload for periods as short as two seconds can be measured with a standard deviation of approximately 1.2 units on a 10 point scale. Figure 2 displays a sample fit between observed and fitted workload values for a single calibration of verbal and spatial workload which included an extra high verbal workload sample.

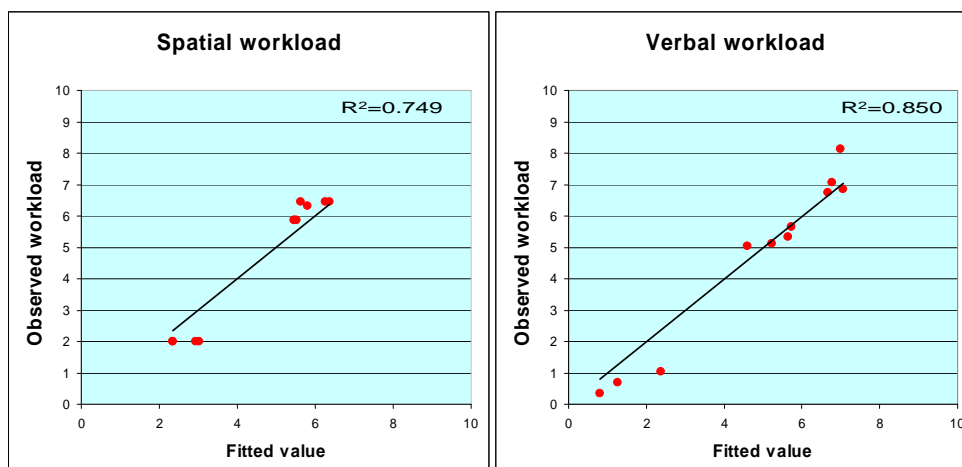


FIGURE 2. Plot of the fit to verbal and spatial workload for a single calibration

The incidence of different frequency components of the EEG in the equations that predicted both spatial and verbal workload were investigated and the summary counts for verbal workload are displayed in Figure 3.

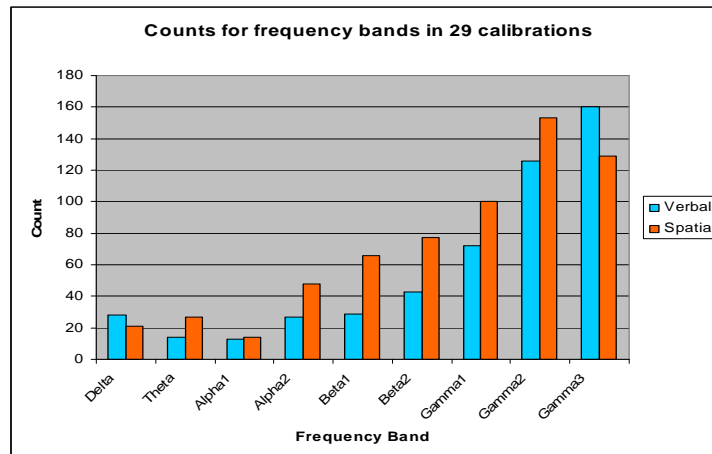


FIGURE 3. The number of components in each frequency band involved in calibration equations

5. Discussion

The results of calibrating the cleaned EEG values to the workload scores for both verbal and spatial workload are clearly quite good. In addition the stated R^2 values include the variation within and between the 4 second blocks that contribute to each 150 second test, although there is a risk that the overall procedure may be fitting to the noise. In the specific case displayed in Figure 2 a single coherence between an electrode pair explained more than half the total variance accounted for in the fit to verbal workload; coherence was low when performing the verbal task and high when it was not performed. This suggests that at least one signal is present for verbal tasks and absent for spatial tasks, providing support to the assumption that different cognitive activities are involved in performing verbal and spatial tasks.

From the results shown in Figure 3 there is a strong tendency for the set of variables that explain variation in workload to be dominated by spectral components with frequency above 30Hz. There is no possibility that mains frequency is the culprit since the band between 47Hz and 53Hz is excluded from the spectral estimates. Since the workload prediction equations are dominated by high frequency components there is clearly a risk that the estimates of gain and coherence may be contaminated by extraneous signals or muscle artefact and the removal of these signals is critical for the success of the procedure. To control for the generation of extraneous signals from display equipment during calibration, both tasks use the same display with similar content, reducing the likelihood that extraneous signals could be different for verbal and spatial tasks. There is a difference in subject motor activity for the Bakan and tracking tasks. In the former case the participant responds relatively infrequently while for the tracking task the participant makes almost continuous movements and it is possible that motor activity could affect the findings. In addition, kurtosis based ICA will be effective for eliminating anomalous observations in a block so long as the contamination is present less than 20% of the time and the required signal is present for the majority of the time. As the contaminated fraction increases or even dominates the recording the model underlying the artefact removal procedure will break down and cleaning will be less successful. The

procedure for marking blocks as unsatisfactory is designed to manage this situation and experience to date indicates that this procedure works reasonably well.

The workload measurement method was originally developed under the DARPA programme for employment in adaptive systems. The deployment of EEG measurement in the field in this role is some way off. The methodology described in this paper can be used in the laboratory as part of the system assessment process in that it provides a continuous measure of both verbal and spatial workload. It can provide valuable insight into the dynamic pattern of task load in future military systems and can support role allocation in time-critical functions as part of the process of ensuring that military effectiveness is maintained under all operating conditions.

REFERENCES

- BAKAN, P. 1959. Extraversion-introversion and improvement in an auditory vigilance task. *Br J Psychol.*, **50**, 325-332
- BELYAVIN, A., RYDER, C., & DICKSON, B. 2007. Development of Gauges for the QinetiQ Cognition Monitor. In *Augmented Cognition HCII 2007*. 3-12.
- BELYAVIN, A.J. & FARMER, E.W. 2006. Modeling the workload and performance of psychomotor tasks. *2006 Conference on Behavior Representation in Modeling and Simulation (BRIMS)-022*. Baltimore.
- JOHNSTONE, I.M. 2001 On The Distribution Of The Largest Eigenvalue In Principal Components Analysis. *The Annals of Statistics*, **29**, 2, 295-327.
- MOULINES, E. & CARDOSO, J. 1992. Direction finding algorithms using fourth-order statistics. Asymptotic performance analysis, In *Proc. ICASSP*, 437-440, March 1992.
- NICHOLLS, A. P., FARMER, E. W., PEACHEY, R. I., & BELYAVIN, A. J. 2003. Dual task interference: Using Hierarchical Cluster Analysis to identify underlying cognitive mechanisms. *Proceedings of the 47th Human Factors and Ergonomics Meeting*. Denver, Colorado, US. Oct 13-17, 2003
- TRACY, C.A. & WIDOM, H. 1994. Level-spacing distributions and the Airy kernel. *Communications in Mathematical Physics* **159** (1): 151-174