EAI Endorsed Transactions
on Context-aware Systems and Applications
Research Article EAI.EU

Temporal assessment of cognitive load factors using ocular features during a visual search

Minoru Nakayama[1,*], Tomomi Okano[1]

[1]Information and Communications Engineering, Institute of Science Tokyo (Tokyo Tech.)
O-okayama, Meguro-ku, Tokyo, 152–8552 Japan

Abstract

The possibility of evaluating temporal changes in cognitive workloads during a visual search task is examined using microsaccade (MS) rates and pupillary changes. The experimental task was designed as a search for a specific figure, where task difficulty and reaction accuracy during the trials were controlled. Individual cognitive workloads were measured after the experimental sessions were conducted, using NASA-TLX scale ratings. Temporal changes in the cognitive load were identified using metrics of oculomotors during two stages of task processing, by comparing cognitive loads with individual ratings on a scale. Since the source of the load may be a common one, changes in latent attention resources required for the task were estimated with a designated state-space model, using the observation data in order to synthesise measurement of MS rates and pupillary changes. The predicted levels of attention resources correspond to the activity during the performance of the experimental tasks during the trials, and reflected some of the rating scores for workload scales. Also, the ranges of confidence intervals for attention resources correlate significantly with the ratings for information processing at the stage where visual stimulus is presented during tasks.

Received on 26 February 2025; accepted on 09 May 2025; published on 05 August 2025

Keywords: cognitive load, microsaccade rate, pupil change, modelling, chronological analysis

doi:10.4108/eetcasa.8797

1. Introduction

Assessment of task difficulty and cognitive load are key factors in the improvement of visual operation performance during human-computer interaction (HCI), and in the usability of HCI systems. In particular, temporal measurements are imperative for evaluation of the operational procedures of these systems. Bio- logical measurements are the metrics often used to observe changes in responses during task operation and performance. Since eye tracking and pupillary metrics of attention or cognitive load can be observed visually, these metrics are frequently used to examine the suitability of user interfaces [1, 2]. As a typical eye tracking metric, microsaccades (MSs) are often referred to as an index of high level cognition [3, 4]. Also, pupil response indicates cognitive activity, and in some studies pupil sizes are measured synchronously using eye tracking measurements [5, 6]. Pupil response has some latency in reacting to the source of influence, and thus simultaneous analysis using other ocular metrics may not be easy [7]. If the same source affects these ocular responses, their relationships and mechanisms should be investigated and analysed [8]. Recent mod- elling techniques may be able to extract latent activity, so that the relationship between task performance, the level of task difficulty, and factors of the visual image presented can be analysed.

In this paper, visual search task performance is analysed using observed temporal MS frequency and pupillary changes, by comparing these with participant's ratings of the level of cognitive load [9, 10]. The changes in ocular metrics and correlation with ratings of cognitive load are evaluated during each stage of the experiment. The contribution of the experimental settings is extracted using a state-space modelling technique, and chronological task activity is monitored [11–13].

*Corresponding author. Email: nakayama@ict.e.titech.ac.jp

The following topics will be addressed in this chapter.

- Changes in ocular metrics in response to the level of cognitive load during two stages of a visual search task are examined.
- Levels of cognitive workload during search tasks is estimated using a state-space modelling technique with microsaccade rates and pupil diameters.
- The temporal changes in workload are analysed using experimental factors which may affect responses.
- Factors of cognitive workload can be extracted based on their comparison with the ratings for the conventional measurement of workloads, which in this case are NASA-TLX scales.

The remainder of the paper is structured as follows. Section 2 reviews related works regarding this topic. Section 3 presents the experimental method, including the task and procedures. Results of the metrics measured are summarised in Section 4. Estimation of latent activity during the task using modelling techniques is discussed in Section 5. An overall discussion is presented in Section 6, and the conclusions of this paper are summarised in Section 7.

2. Related works

2.1. Cognitive load and ocular metrics

Generally, human information processing requires attention or cognitive load to perform any task. If the workload amount is above the processing capacity, this may lead to the failure of the task or its incomplete processing, such as when accidents occur while driving if motorists miss visual objects on the road like obstacles or potholes.

However, most assessments are based on subjective human reports or on psychometric scales. One of the major scales is the National Aeronautics and Space Administration Task Load Index (NASA-TLX) [14]. This scale has been developed for various tasks by adjusting the weights of 6 of the factor ratings of the index [15, 16]. This technique is intended to evaluate the usability of a developed system, using specialised assessment techniques developed for this purpose [2]. In these applications, ocular metrics are also employed [1]. As ocular metrics can be measured temporally, chronological assessment regarding task processing may be possible [17–19]. Therefore, the metrics of microsaccades (MSs) and pupil response have often been used [4, 20]. In particular, MS frequency and MS characteristics are often used to analyse high level information processing, such as with indicators of the covert orienting of attention [21, 22].

While these metrics may respond to cognitive load, they are not well synchronised [5], and cause response latencies as a result. The two metrics respond to cognitive load factors, but the conditions or the characteristics of responses may be different [23]. Therefore, overall cognitive load or attention levels should be estimated using these measurements.

2.2. Modelling approach

Biological responses can be observed as peripheral reactions, and include ocular activities. Cognitive load may occur as a form of latent behaviour. Even if observed metrics show different tendencies, they may be due to common factors such as the latent behaviour. As the latent activity can not be measured directly, an estimation technique may aid in the recognition of the chronological change.

Recently, the Bayesian inference approach has provided some advantages to obtaining solutions to phenomena having insufficient data, such as where only a limited number of measurements are available [24–26]. There are also several calculation platforms which allow the introduction of various types of data [27–29]. In regards to this approach, the conventional experimental paradigms have been reanalysed and new evidence of mental mechanisms has been extracted [30, 31].

Temporal changes during the observation of eye tracking, including eye movement and pupillary changes, can be analysed using this approach [11, 32].

3. Experimental Method

Oculo-motor metrics such as eye movement and pupil diameter were measured during a visual search task designed for this purpose [9, 10]. A presentation diagram is shown in Figure 1. A visual search task is presented after the fixation of the eyes on the centre of the display ("Ready" stage). During a task (the "Go- task" stage), a set of figures was presented. The visual search task consisted of the counting of a targeted shape in a series of images which contained 7 different kinds of line-drawing figures, such as circles, triangles, squares, pentagons, and other shapes.

3.1. Procedure

The stimuli of line drawn figures were presented in the centre of a 87mm circle (visual angle: 10 deg.) with a white background, and presented on a 27-inch LCD screen (Eizo, EV2736WZ) which was positioned 530mm away from the observer. Since the stimuli were illustrated using line drawn shapes, the level of luminance of the display background and the brightness of room remained constant, even as the number of shapes increased.

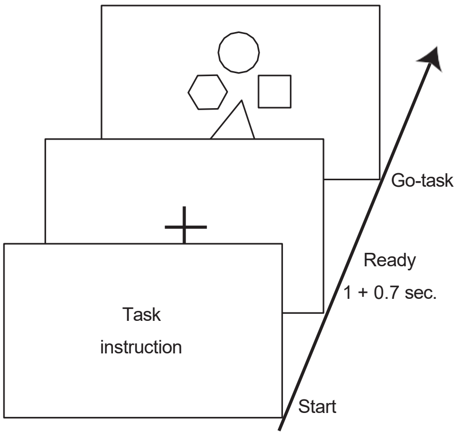

Figure 1. Sequence of experimental stimuli

Experimental stimuli were presented in the following order, as shown in Figure 1.

1. An instruction regarding the subsequent trial was presented
2. Fixation on a central point for 1.0 second
3. A blank image or "refresh" was displayed for 0.7 seconds (not illustrated)
4. A stimulus was presented and 3 alternative responses were provided

An experimental program was created using MATLAB and Psychtoolbox-3 and presented to participants of the experiment.

Each participant completed a set of 20 tasks twice (20 tasks × 2 sets, for a total of 40 tasks) with a short break in between sets so that participants could refresh themselves. Some participants took part in their two sessions on different days.

The participants were 10 undergraduate university or graduate students (5 male and 5 female with a mean age of 21.9 years [SD=0.85]) who possessed sufficient visual acuity (both eyes above 0.8 [styled as 20/25 in some countries] with the naked eye or with corrected vision). Prior to the experiment, a complete explanation of the content of the experiment was given and formal consent was obtained (Institutional approval: #2019052).

3.2. Experimental conditions

The "bottom-up factor" for eye movement as a feature of visual stimuli displayed was evaluated using the following metrics.

- Task difficulty and rate of correct responses

  The visible size of stimuli decreased as the trial sequence proceeded, and the increase in task difficulty of detecting the target was controlled as a part of the design of the experimental. The responses to stimuli were classified into two levels of difficulty based on correct reaction times and cluster analysis [9].

- Saliency of visual stimuli

  Since the quantity of figures presented increased during the trials, visual complexity also increased gradually, along with task difficulty. Complexity was evaluated using the saliency of the visual image stimuli, and saliency was computed using OpenCV and StaticSaliency [10, 33].

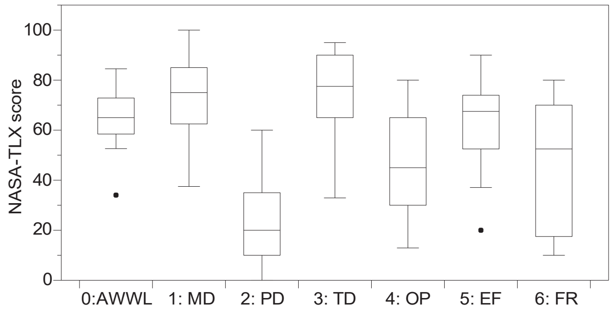

Figure 2. Scores of NASA-TLX

3.3. Measurement of ocular metrics

Microsaccade rate and pupil diameter. Both eye movement and pupil diameters were tracked using a ViewPoint EyeTracker (ArringtonResearch: BCU400, 400Hz) with a chin rest. Microsaccade rates were extracted using MS ToolBox [21, 34, 35]. Pupil sizes of each participant were standardised using a mean pupil sizes measured during the 1.2s before stimuli were shown. Also, pupil diameters during blink were removed, resulting in a vertical/horizontal aspect ratio of less than 0.7.

Cognitive workload. The workload of each participant was evaluated using a visual analogue scale (VAS) on the NASA-TLX [15, 16]. The NASA-TLX is commonly used as a subjective assessment scale to measure the cognitive load of various workloads. In order to adapt this to a specialised task, the overall score was calculated using appropriated weights of the following 6 factors [14]. In this work, the inverted rank order of mean ratings was used. This consisted of 6 factors: Mental demand(MD), Physical Demand (PD), Time Demand (TD), Performance (OP), Effort (EF) and Frustration (FR). Overall workload was estimated as AWWL (Adaptive Weighted Workload) using 6 scores and their weights [14, 36].

4. Results of Responses

4.1. Participant assessment of own cognitive load

For each session the cognitive workload during the experimental task was measured using 6 dimensional

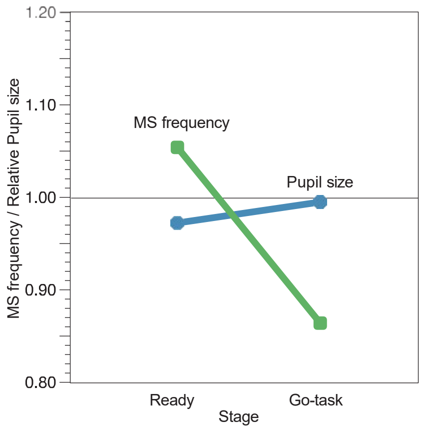

Figure 3. MS frequency and Relative Pupil size during two stages.

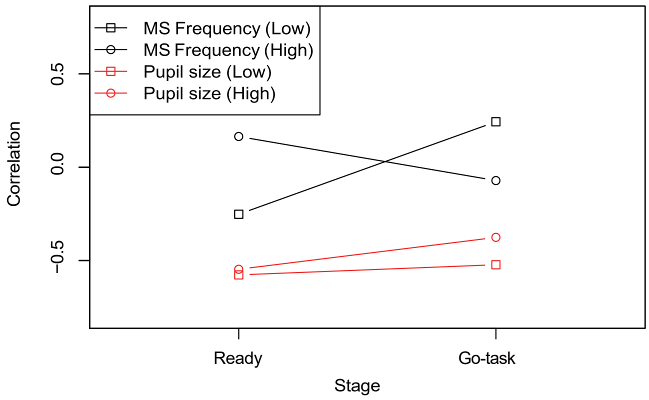

Figure 4. Correlation of features of oculomotors with Performance rating (OP).

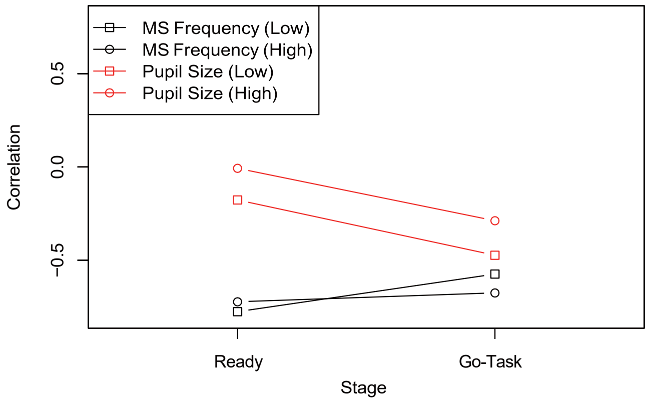

Figure 5. Correlation of features of oculomotors with Frustration rating (FR).

NASA-TLX scales. The rated responses are summarised as box plots in Figure 2. The scores of "0: AWWL" on the left side of the figure show the results of the overall degree of cognitive load, which is based on the other 6 item ratings of the 20 trials in total (2 trials × 10 subjects). Ratings for mental demand (MD) and time demand (TD) are higher than for the other items, and physical demand (PD) is rated the lowest. These results confirm that the high requirement for mental demand (MD), time demand (TD) and effort (EF) represent aspects of the experimental task.

4.2. Temporal changes in ocular metrics and levels of cognitive load

Overall means of MS frequency and relative pupil size during two experimental stages are summarised in Figure 3. MS frequency was suppressed significantly between the "Go-task" stage and the "Ready" stage ($t(4626) = 3.64$, $p < 0.01$; Cohen's d=0.10), while pupil size increased significantly between the "Ready" and the "Go-task" stages ($t(4130) = 5.89$, $p < 0.01$; Cohen's d=0.17). These results show that the workload necessary for completing tasks affects both ocular metrics.

MS Frequency and Pupil size. In order to extract factors of the changes in metrics, correlation analysis was conducted using NASA-TLX measurements in order to the evaluate contribution of participant's perceptive workload during the task. Correlation analysis was conducted to assess workload levels at two levels of task difficulty using the above mentioned extracted data. Two typical cases are summarised. In the first case, the correlation coefficients of the Performance rating (OP) and the ocular metrics of the two stages are summarised in Figure 4. Pupil sizes remain correlated with the OP ratings of both stages, though coefficients of MS frequency remain around 0. There is little difference in the difficulty of the task. In the other case, correlation coefficients of the Frustration rating (FR) in Figure 5 are summarised. In this figure, MS frequency remains correlated with the FR ratings of both stages, while pupil sizes correlate with FR ratings in the "Go-task" stage. These results suggest that the two ocular metrics may reflect a specific aspect of the workload.

Correlation analysis. Correlation analysis was conducted on all ratings of cognitive load factors of the two levels of task difficulty.

Correlation coefficients for Low task difficulty are summarised in Table 1, and the values for High task difficulty are summarised in Table 2. In these tables, only significant coefficients are displayed. As the tables show, pupil size during the "Go-task" stage correlates with some of the factors of workload even at a low level of task difficulty. As all coefficients are negative, smaller pupil sizes affect the higher ratings of workload factors. For high levels of task difficulty, some of the factor ratings correlate with MS frequency. Factor ratings of frustration (FR) correlate with MS frequency at both levels of task difficulty, as shown in Figure 5.

Table 1. Correlation between NASA-TLX Items and Microsaccade Frequency, Pupil Diameter. (Low Difficulty)(N=20)

	MS frequency		Pupil size	
Items	Ready	Go-task	Ready	Go-task
AWWL	-	-	-	-0.45
TD	-	-	-0.36	-0.39
OP	-	-	-0.58	-0.52
FR	-0.78	-0.57	-	-0.53

Table 2. Correlation Coefficient between NASA-TLX Items and Microsaccade Frequency, Pupil Diameter. (High Difficulty)(N=20)

	MS frequency		Pupil size	
Items	Ready	Go-task	Ready	Go-task
MD	-0.36	-0.52	-	-0.36
PD	-	0.40	-	-
OP	-	-	-0.55	-0.38
FR	-0.72	-0.67	-0.33	-0.54

4.3. Discussion

Ocular metrics of MS frequency and pupillary changes were measured in response to visual search task which controlled the difficulty of the task in trials. Participant's overall cognitive workload was measured using a rating scale of 6 factors of cognitive load based on NASA-TLX. In order to focus on the change of experimental stages, such as from the "Ready" stage while awaiting stimuli to the "Go-task" stage for the response to the visual search task, the measured metrics were compared.

Both metrics of MS frequency and pupil size changed significantly between the two stages, since the experimental task affects ocular activity in response to the cognitive load. The contribution of these activities on the ratings of measured perceptive cognitive load was analysed using correlation analysis. There are some significant correlation relationships between the two levels of task difficulty. Both ocular metrics correlate with some of the factor ratings, however the correlation relationship between the two metrics is different. Therefore, the two metrics may represent different aspects of cognitive load.

In response to the cognitive load of the experimental tasks, 6 factors are extracted for use as a measurement scale [14]. Cognitive workload levels are then extracted from the metrics using these scales, which is the central purpose of this work. As the experimental stages are not independent activities, the cause of the changes should also be explained. A new analytical approach will be introduced in the next section in order to extract and analyse the metrics of this temporal processing activity.

5. Modelling with microsaccade rate and pupil size

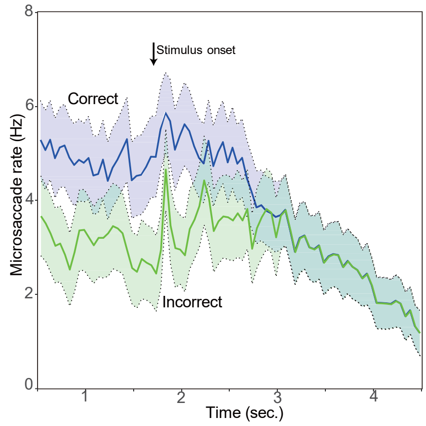

Figure 6. Temporal change in MS rates for correct and incorrect responses.

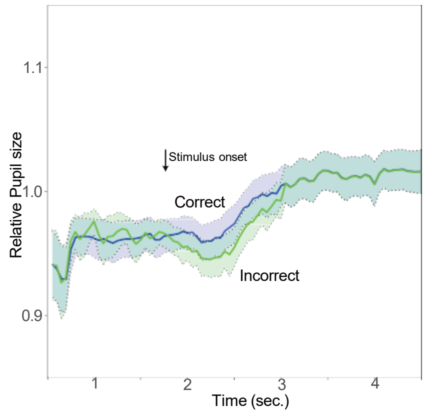

Figure 7. Pupillary change for correct and incorrect responses.

5.1. Observed data

A modelling technique used for estimating attention levels in a previous study of ours [11] is applied to the attention level measurements using a state-space model based on both the MS rate and pupil size. Both metrics were measured continuously during the experimental session. The observation period is focused on the 0.5∼4.5 second duration of each trial in regards to the distribution of a participant's reaction time. Mean reaction time for the task was 2.75 seconds after stimulus onset. Mean metrics for MS rates during this period are summarised in Figure 6 and for relative pupil size in Figure 7. In these figures, mean temporal changes with a 95% confidence interval between correct and incorrect responses are compared. There are some differences between them, but after 3 seconds they become synchronised.

5.2. Model structure

The model is designed as follows [27, 28]. The estimated "Attention resource (Attn)" of each participant isdefined in equation (1) as a summation of the level of attention during a task (*S_level*) as noted in equation (2), and the intercept for each consists of the response correctness (*Correctness* in two dimensions: correct and incorrect), the task difficulty (*T askD* in two dimensions: High and Low), individual factors (*rID* times 2 for 10 participants: 20 dimensions) and the stimulus order factor (*rPN* in 20 dimensions). As a hypothesis in the observation model, the change in observed MS rates is noted using a Poisson distribution because of the frequency of the occurrences, and the change in pupil size around an average is noted using a Normal distribution and its deviation [11, 12]. The tendency for pupillary changes in response to the level of attention is the opposite of the change in MS rate, as inverse transform is applied to the pupil responses [11].

State Model:

$$Attn = S\_level + Correctness + TaskD + rID + rPN \tag{1}$$

$$S\_level_i \sim Normal(S\_level_{i-1}, \sigma_s) \tag{2}$$

Observation Model:

$$mu_{noise} \sim Normal(Attn, \sigma_{noise})$$

$$\lambda = exp(\mu_{noise})$$

$$MS_{times} \sim Poisson(\lambda)$$

$$Pupil_{size} \sim Normal(Attn, \sigma_p)$$

All model parameters were estimated with the observation data and random sampling values using the MCMC (Markov Chain Monte Carlo) technique, which generates random values (4000 iterations with 4 chains and a burn-in of 500). Conversion of model parameters is confirmed as $\hat{R} < 1.1$, with the number of states controlled to extract optimum conditions such as $(i = 1 \ldots N, N = 4, \ldots, 16)$ using R and Stan packages. A possible conversion solution shows that the model can estimate a latent level using the estimated parameters and the observation data. From the optimisation results, the number of states is set at 16. In addition, individual factors (*rID*) were assigned to different participants in the 1st and 2nd sessions because some participants participated as two separate individuals in two sessions.

5.3. Results

Estimated parameters. Estimated parameters are summarised in order to evaluate the model. Common latent attention levels during tasks can be represented in equation (2) as a series of parameters (*S_level*), which are sources of attention changes in the previously mentioned equation (1). These posterior distributions are illustrated in Figure 8. The horizontal axis represents time bins (1∼16) in 0.25 second/bin increments. The vertical axis represents attention activity. The red band

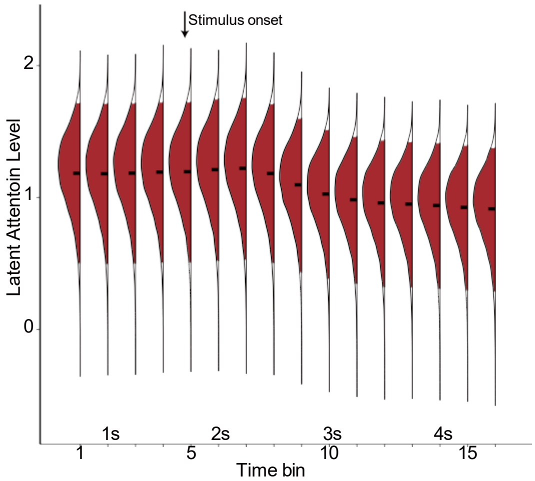

Figure 8. Latent attention levels

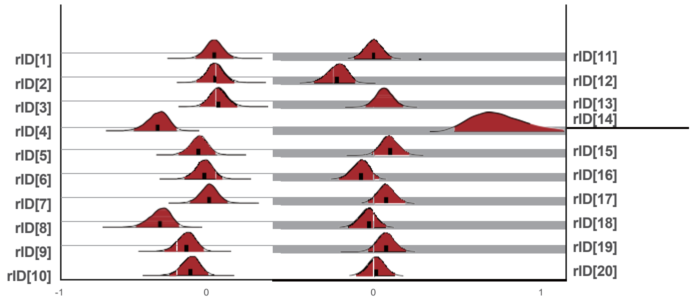

Figure 9. Individual factors of experiment participants

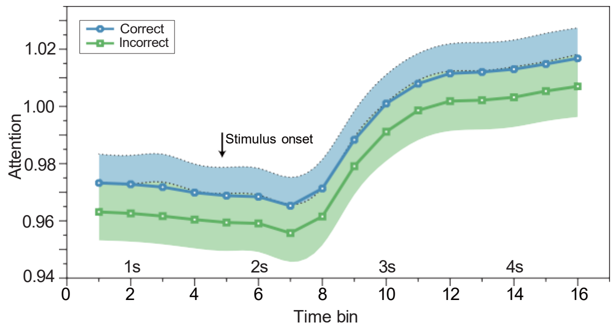

Figure 10. Attention levels of correct and incorrect responses.

represents a confidence interval of 95%. The highest level of estimation is at 2.25 seconds (time bin=7) after stimulus onset at 1.7 seconds. The level of estimation decreases from its highest (2.25 seconds) to its lowest around 4 seconds. An additional estimated parameter of individual factors (*rID*) is summarised in Figure 9. Participants are divided into two columns. A pair of horizontally displayed distributions of (*rID*) from two trials by the very same participant show deviations, which means that some participants produced different distributions, such as for "rID[4]" and "rID[14]", during each of their sessions. Therefore, individual factors of

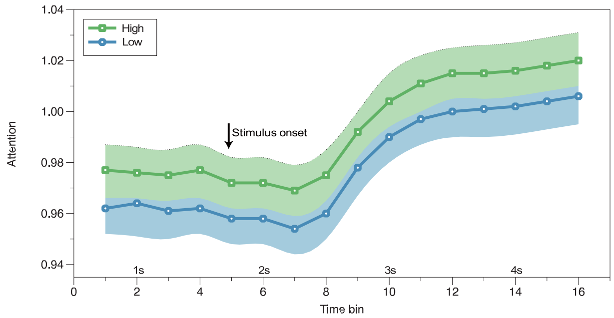

Figure 11. Comparison of saliency (L,H) of visual stimuli

all who participated in the two sessions are designated as factors of separate individuals. This is the reason why participants measurements are recorded separately in each of their two sessions, as mentioned above.

The estimated attention resource ($Attn$) for correct and incorrect responses which are affected by another parameter ($Correctness$) is illustrated in Figure 10, where the blue line represents cases which are correct, and the green line represents cases which are incorrect, with a 95% confidence interval. Overall, the estimated levels ($Attn$) for correct responses are significantly higher than the levels for incorrect responses ($p < 0.05$, Cohen $d >$ 0.24), though the value of the difference in ($Correctness$) is small. For incorrect responses, attention resources are more suppressed due to the requirement to perform complicated task processing. The overall trend shows that the level of attention resources is the smallest at 2.25 seconds, time bin=7, 0.5 seconds after stimulus onset at 1.7 seconds. The differences between the two responses are larger around just after stimulus onset (1.7 ∼ 2.45 seconds; $p < 0.01$).

Another estimated parameter of task difficulty ($T askD$) also affects attention resources ($Attn$), and their temporal changes are similar to the results of response correctness in Figure 10.

When the contribution of individual factors to attention resources is considered, the factor for correct response is not significant. The factor for task difficulty at 1.95 seconds just after stimulus onset (1.7 seconds) contributes to the decrease of the resource. The task difficulty may affect the change in attention resources.

Contribution of saliency of visual stimuli. In addition to comparing the levels of attention resources of the estimated parameters, the contribution of image saliency of visual stimuli is also examined. Values of image saliency for 20 visual stimuli are calculated, and classified into two levels (High and Low). Temporal changes in levels of attention resources are illustrated in Figure 11. The minimum resource level at 2.25 seconds (time bin=7) after stimulus onset also shows that the resources are being used for task processing. The suppression of resources decreases between the Low and High levels. When the level of image saliency is high, it may be easier to perceive visual stimulus, as suppression is reduced. On the other hand, low image saliency requires more attention, so the level of resources is wasted or reduced. These changes in response to levels of image saliency and to the order of the level of attention resources were also confirmed, when three levels of image saliency of visual stimuli are classified.

This result shows that even attention resources ($Attn$) respond to external parameters of the model equations such as image saliency.

Impact on cognitive workload. Since the level of estimated attention resources may affect individual NASA- TLX scale ratings, these correlation relationships are examined. The correlation coefficients of the estimated parameters and the ratings for NASA-TLX are summarised in Table 3. The horizontal column represents NASA-TLX scales, and the coefficients for averaged attention resources over trials ($Attn$), individual factor (rID), and range of confidence interval for attention resources (CId) are calculated. Some significant coefficients for attention resources ($Attn$) are observed in columns for time demand (TD), effort (EF), and overall value (AWWL). However, because most coefficients are positive values, this seems a little bit strange as more attention resources correlates with a higher task workload. Temporal changes in attention resources show a similar tendency. The attention resources ($Attn$) are given as a summation of several factor values shown in the equation (1). Though correlation coefficients are always positive during temporal changes, the parameter known as attention level (S_level) may not contribute to the relationship.

As the ratings for task workload are the responses of individuals, the contribution of individual factor (rID) is evaluated as a coefficient on the second line of Table 3. Most coefficients are negative, and absolute values are comparable with coefficients for attention resources ($Attn$). The individual factor (rID) is an intercept for the equation (1), and it adjusts the change in attention level (S_level) of each participant. Therefore, the individual factor (rID) may represent a baseline for attention resources during a task, so it can constitute the cognitive workload of each individual.

In addition, temporal changes in attention resources show that the range of the confidence intervals is minimised during a period of time after stimulus onset. The deviation of confident intervals may change with the activity of task processing. Then, the range of the confidence intervals across trials (CId) is defined in order to examine the relationship between the level of workload and the change in attention resources ($Attn$). Correlation coefficients between the range of

Table 3. Correlation coefficients between NASA-TLX scores and estimated parameters

est.	AWWL	MD	PD	TD	OP	EF	FR
Attn	0.55	(0.30)	(-.05)	0.49	(0.17)	0.46	(0.33)
rID	-.60	(-.36)	(0.03)	-.56	(-.16)	-.46	(-.35)
CId	(-.23)	[-.43]	(-.15)	(-.33)	(-.05)	(0.01)	(0.18)

N=20 (10 participants × 2 sessions)

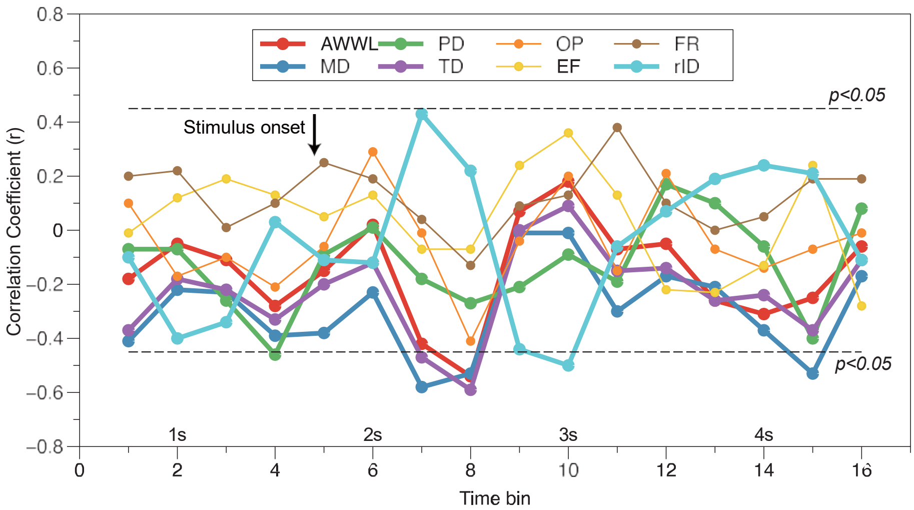

Figure 12. Correlation coefficients between NASA-TLX scores and ranges of confidence intervals

the confidence interval (*CId*) and ratings of NASA-TLX scales are summarised on the third line of Table 3, however the absolute values are relatively small and are not significant.

For further analyses, calculation of the correlation coefficient is extended chronologically to every time bin. Though overall means of the range of the confidence interval (*CId*) do not contribute to the NASA-TLX scale ratings, they react to certain time bins. The correlation coefficients of their relationships using the 6 factor scales and AWWL are summarised chronologically in Figure 12. The horizontal axis represents time bins, and the vertical axis represents correlation coefficients. Two horizontal dotted lines show significant levels of probability for the coefficient ($p < 0.05$). The overall trends of these changes seem similar. Negative coefficients for mental demand (MD) and time demand (TD) are significant between 2.20 and 2.45 seconds. This suggests that the range (*CId*) during these periods is reduced when the ratings are large. Also, the coefficient for mental demand (MD) is significant at a time bin of 4.25 seconds, which is as around the mean reaction time for the task. The coefficient may reflect the individual's task response times.

Another coefficient for individual factor (*rID*) changes contrary to the ranges of other values. In particular, the coefficient shows the highest at time bin 7 after stimulus onset and the lowest at time bin 10 in response to change in attention levels (*S_level*).

These results present the possibility that temporal changes in the estimated attention resources (*Attn*) can be used for chronological assessment of the cognitive workload.

6. General Discussion

When correlation analysis was introduced to examine the relationships between each ocular metric using the experimental stages and individual ratings of cognitive load, the results were summarised but were inconsistent, though some significant differences were observed. These results suggest that every metric may detect different aspects of latent activity and their temporal changes. These ocular metrics provide information about the possibility of evaluating human behaviour and the effect of experimental conditions on visual search tasks. However, it is not easy to understand well nor summarise the changes in cognitive load using the results analysed.

A modelling technique was introduced in order to summarise the responses and evaluate the factors of the experimental conditions. A modelling technique using individual MS rates and pupillary changes is introduced in order to estimate the amount of attention resources required during task processing.

The estimated results show that attention resources are reduced by the workload produced when the task has not been performed completely. The change in attention resources ($Attn$) is caused by variation of MS rates and pupillary changes, in addition to experimental factors such as parameters for $Correctness$ and $TaskD$ in equation (1). So, temporal changes in the estimated attention resources ($Attn$) can be recognised as the amount of attention resources provided in order to perform the task.

In regards to the change in MS rates in Figure 6, the MS rate is reduced after the release of cognitive attention and the dropping off of the stimulus onset. Since there are no significant differences in MS rates between correct and incorrect responses after 3 seconds, the task seems to be completed by around 3 seconds. Also, some differences in pupillary changes can be observed between stimulus onset and the 3 second point in Figure 7. Attention activity in Figure 8 suggests a decrease in the level in response to these reactions. In considering the mean reaction time of 2.75 seconds, the change in attention resources ($Attn$) may represent a processing stimulus during the initial stage of perceiving performance. The processing during later stages may consist of managing key input selections.

Figures 10 and 11 show the differences in experimental conditions before stimulus onset. As Figure 8 indicates, attention workload is boosted toward stimulus onset as is clearly presented by the MS rate shown in Figure 6. Since task difficulty was set to increase gradually, the rate of correct responses and reaction time became worse during the second half of the experimental session. There is the possibility that a previous trial might have influenced the beginning of the next task. However, the estimated parameter for the factor of stimulus order is not influenced by the presentation order. Pupil response shows a simple difference during task performance. Though the actual factor cannot be determined, these composite factors may influence the differences in metrics before stimulus onset.

The estimated attention resources may reflect a level of cognitive workload, because both metrics are influenced by workload during the performance of the task. As an overall assessment, mean attention resources ($Attn$) correlate significantly with ratings for time demand (TD), effort (EF) and the summation (AWWL) of NASA-TLX scales. In particular, the individual factor (rID) is a main component of the attention resources and contributes to the correlation relationship, as shown in Table 3. For a chronological analysis, the range of confidence intervals for attention resources (CId) can serve as a mean for measuring changes in the cognitive workload level. The range of the confidence interval represents the degree of deviation in the estimated attention resources, which consist of individual responses in MS rate and pupillary responses. This means that instability of attention resources reflects levels of cognitive workload. Though the remaining parameters in equation (1) are constant during the experimental task, indicated levels of responses to the workload vary over the time course. The temporal changes in correlation coefficients in Figure 12 show workloads during the progress of performing a task.

The parameters analysed are estimated using a state-space modelling technique which extracts the latent activity necessary in order to perform a task. The results were based on individual responses and adapted them. As a few differences between individual responses during the two sessions were evident, some experimental conditions and tasks may affect the ocular reactions. These results are based on the specific conditions of the experiment. Also, all assessments are based on post modelling of the experimental measurements. Development of real time analysis and feedback regarding latent information processing will be subjects of our further study.

7. Conclusion

Ocular information such as MS rates and pupillary changes were observed in order to evaluate cognitive load and experimental conditions during the performance of a visual search. Participant's task reactions were analysed by comparing ocular information with participant's ratings of the cognitive load using NASA-TLX scales. The responses, which included ocular information and experimental factors, were evaluated separately and their contribution to the cognitive workload was discussed. Also, a state-space model which represents latent attention resources in response to solving the experimental tasks was introduced in order to conduct an overall assessment.

The following results are summarised.

1. Both MS frequency and pupillary changes were summarised in response to the task difficulty and experimental conditions. The relationship between ocular metrics and individual ratings of NASA-TLX scales were analysed. Some significant correlation relationships were observed, and the possibility of detecting specific cognitive load was examined.

2. All parameters of the state-space model which was developed were estimated using MS rates and pupillary changes measured during the experiment. The number of states was optimised.

3. Attention resources were estimated using a model, and the estimated base value measurements correspond with the conditions of the experiment and the progress of task performance.

4. The estimated attention resources represent some ratings of cognitive workload, such as time demand, effort and summation of ratings. In particular, individual factors directly illustrate the level of the workload, and individual ranges of attention resources correlate chronologically with some workload ratings in response to the progress of the performance of the task.

Also, the problems which remain are summarised for further study.

Acknowledgement

This research was partially supported by the Japan Society for the Promotion of Science (JSPS), Grant-in- Aid for Scientific Research. (KAKEN, 20H01718: 2020- 2023).

References

[1] Cairns, P. and Cox, A.L. (2008) *Research Methods for Human-Computer Interaction* (Cambridge, UK: Cambridge University Press).

[2] Preece, J., Rogers, Y. and Sharp, H. (2015) *Interaction Design beyond human-computer interaction, 4th edition* (Chichester, UK: John Wiley & Sons Ltd.).

[3] Kohama, T., Nakai, Y., Ohtani, S., Yamamoto, M., Ueda, S. and Yoshida, H. (2017) Quantitative comparison of cognitive load derived from voice or manual responses based on microsaccade rate analysis. *Journal of Human Interface Society* 19(2): 189–198.

[4] Dalmaso, M., Castelli, L., Scatturin, P. and Galfano, G. (2017) Working memory load modulates microsaccade rate. *Journal of Vision* 17(3): 1–12.

[5] Krejtz, K., Duchowski, A.T., Niedzielska, A., Biele, C. and Krejtz, I. (2018) Eye tracking cognitive load using pupil diameter and microsaccades with fixed gaze. *PloS One* 13: 1–23.

[6] Dalmaso, M., Castelli, L. and Galfano, G. (2020) Microsaccadic rate and pupil size dynamics in pro-/anti-saccade preparation: the impact of intermixed vs. blocked trial administration. *Psychological Research* 84: 1320–1332.

[7] Klein, C. and Ettinger, U. (2019) *Eye Movement Research, An Introduction to its Scientific Foundations and Applications* (Springer Nature Switzerland AG).

[8] Nakayama, R., Bardin, J.B., Koizumi, A., Motoyoshi, I. and Amano, K. (2022) Building a decoder of perceptual decisions from microsaccades and pupil size. *Frontiers in Psychology* 13: 942859.

[9] Okano, T. and Nakayama, M. (2021) Feasibility of evaluating temporal changes in cognitive load factors using ocular features. In *COGAIN workshop, Proceedings of ETRA21*: 29:1–6. https://doi.org/10.1145/3448018.3458019.

[10] Okano, T. and Nakayama, M. (2022) Research on time series evaluation of cognitive load factors using features of eye movement. In *COGAIN workshop, Proceedings of ETRA22*: 61:1–6. https://doi.org/10.1145/3448018.3458019.

[11] Dubiel, M., Nakayama, M. and Wang, X. (2023) Modelling attention levels with ocular responses in a speech-in-noise recall task. In *Proceedings of ETRA2023* (ACM): 89:1–7. https://doi.org/10.1145/3588015.3589665.

[12] Nakayama, M. and Ueno, T. (2023) Estimation of latent attention resources using microsaccade frequency during a dual task. In *Proceedings of ETRA2023* (ACM): 41:1–2. https://doi.org/10.1145/3588015.3590120.

[13] Nakayama, M. and Okano, T. (2024) Extracting cognitive workload factors on modelling attention resources using ocular information in figure search task. In *Proceedings of 28th International Conference Information Visualisation iV*: 95–100. https://doi.org/10.1109/IV64223.2024.00026.

[14] Hart, S.G. (2006) NASA-task load index (NASA-TLX); 20 years later. In *Proceedings of the human factors and Ergonomics Society 50th Annual meeting*: 904–908.

[15] Miyake, S. (2015) Special issues no. 3: Measurement technique for ergonomics, section 3:psychological measurements and analyses (6), mental workload assessment and analysis -a reconsideration of the NASA- TLX-(Japanese). *Ningen Kougaku (The Japanese journal of ergonomics)* 51(6).

[16] Miyake, S. and Kuma-Shiro, M. (1993) Subjective mental workload assessment technique-an introduction to NASA-TLX and SWAT and a proposal of simple scoring methods- (Japanese). *Ningen Kougaku (The Japanese journal of ergonomics)* 29(6).

[17] Haruki Mizushina, Kiyomi Sakamoto, H.K. (2011) Relationship between psychological stress induced by workload and dynamic characteristics of saccadic eye movements during task execution. *Transactions of the Institute of Electronics, Information and Communication Engineers Vol.j94-D No.10 pp.1640-1651* .

[18] Ziv, G. (2016) Gaze behavior and visual attention: A review of eye tracking studies in aviation. *The International Journal of Aviation Psychology* 26(3–4): 75–104.

[19] Peiß, S., Wickens, C.D. and Baruah, R. (2018) Eye-tracking measures in aviation: A selective literature review. *The International Journal of Aerospace Psychology* 28(Issue 3–4): 98–112.

[20] Duchowski, A.T., Krejtz, K., Krejtz, I., Biele, C., Niedzielska, A., Kiefer, P., Raubal, M. *et al.* (2018) The index of pupillary activity: Measuring cognitive load vis- à-vis task difficulty with pupil oscillation.

[21] Engbert, R. and Kliegl, R. (2003) Microsaccades uncover the orientation of covert attention. *Vision Research* 43: 1035–1045.

[22] Kashihara, K., Okanoya, K. and Kawai, N. (2014) Emotional attention modulates microsaccadic rate and direction. *Psychological Research* 78: 166–179.

[23] Nakayama, M. and Hayakawa, Y. (2021) Influence of task evoked mental workloads on oculo-motor indices and their connections. *EAI Trans. Context-aware Systems and Applications* 7(23): e2:1–10.

[24] Wagenmakers, E.J., Marsman, M., Jamil, T., Ly, A., Verhagen, J., Love, J., Selker, R. *et al.* (2018) Bayesian inference for psychology. part i: Theoretical advantages and practical ramifications. *Psychonomic Bulletin & Review* 25: 35–57.

[25] Dienes, Z. and Mclatchie, N. (2018) Four reasons to prefer Bayesian analyses over significance testing. *Psychonomic Bulletin & Review* 25: 207–218.

[26] Nguyen, M.H., La, V.P., Le, T.T. and Vuong, Q.H. (2022) Introduction to Bayesian mindsponge framework analytics: An innovative method for social and psychological research. *MethodsX* 9(101808): 1–16.

[27] Lee, M.D. (2011) How cognitive modeling can benefit from hierarchical Bayesian models. *Journal of Mathematical Psychology* 55: 1–7.

[28] Haaf, J.M. and Rounder, J.N. (2017) Developing constraint in Bayesian mixed models. *Psychological Methods* 22: 779–798.

[29] Schönbrodt, F.D. and Wagenmakers, E.J. (2018) Bayes factor design analysis: Planning for compelling evidence. *Psychonomic Bulletin & Review* 25: 128–142.

[30] Muto, H. and Nagai, M. (2020) Mental rotation of cubes with a snake face: The role of the human-body analogy revisited. *Visual Cognition* 28: 106–111.

[31] Muto, H. (2021) Evidence for mixed processes in normal/mirror discrimination of rotated letters: A Bayesian model comparison between single- and mixed-distribution models. *Japanese Psychological Research* 63(3): 190–202.

[32] Nakayama, M. and Ueno, T. (2023) Latent attention resource estimation of peripheral visual stimuli using microsaccade frequency modellin. In *Proceedings of 27th International Conference on Information Visualisation (iV)* (IEEE): 142–147.

[33] OpenCV (2017), Staticsaliency. https://docs.opencv.org/4.x/d5/d87/classcv_1_1saliency_1_1StaticSaliency.html.

[34] Engbert, R. (2006) Microsaccades: a microcosm for research on oculomotor control, attention, and visual perception. *Progress in Brain Research* 154: 177–192.

[35] Engbert, R., Sinn, P., Mergenthaler, K. and Trukenbrod, H. (2015) Microsaccade toolbox 0.9. *Potsdam Mind Research Repository* .

[36] Haga, S. and Mizugami, N. (1996) Mental workload measurement using NASA-TLX (Japanese version) - sensitivity of workload scores to difficulty of various laboratory experiments-. *Human Engineering 32(2), 71-79* .