Effect of Time Length of Eye Movement Data Analysis on the Accuracy of Mental Workload Estimation During Automobile Driving

Takanori Chihara(✉) and Jiro Sakamoto

Kanazawa University, Ishikawa, Japan
chihara@staff.kanazawa-u.ac.jp

Abstract. We investigated the appropriate time window duration for calculating eye and head movement parameters in mental workload (MWL) estimation during automobile driving. Participants performed driving tasks on a driving simulator, and eye and head movements were measured by controlling their MWL using the N-back task, which required them to keep answering aloud the N-th previous digit in a sequence of digits. The eye and head movement parameters were calculated by changing a time window from 30 s to 150 s in increments of 30 s. An anomaly detector of MWL was constructed using the one-class support vector machine (OCSVM) with the no N-back task ("None") data. In each window length condition, we calculated the area under curve (AUC) for the binary classification between None and the highest MWL condition, the percentage of anomaly data, and the distance from the decision boundary. The results showed that a time window of 30 s had significantly lower AUC compared with other time windows. In addition, the correlation coefficient between the subjective MWL score and the distance of each eye movement parameter data from the decision boundary monotonically increased in the time window 30 s to 120 s and decreased at 150 s. Therefore, we concluded that 60 s to 120 s is an appropriate time window duration for MWL evaluation.

Keywords: Driver monitoring · Mental workload · Eye tracking · Machine learning · Anomaly detection · One-class support vector machine

1 Introduction

Distracted driving is one of the most common causes of fatal traffic accidents in Japan [1]. To prevent distracted driving, it is necessary to evaluate the mental state of drivers. To this end, we focused on the eye movement of the drivers for a quantitative evaluation of mental workload (MWL); tracking eye movements is practical and may indicate comfort, stress, and various other biological states of the driver. In our previous research, the following effective eye movement parameters were selected: the standard deviation (SD) of the horizontal gaze angle, the SD of the horizontal eyeball rotation angle, the sharing rate of head movement against the eyeball rotation in a gaze movement, and the brink frequency [2]. These parameters were calculated in a time window of 60 s, which was determined

N. L. Black et al. (Eds.): IEA 2021, LNNS 221, pp. 593–599, 2021.
https://doi.org/10.1007/978-3-030-74608-7_72

intuitively. However, the duration of the time window may affect the accuracy of MWL estimation. A longer time window may improve the estimation accuracy as the effect of the surrounding noise is reduced; however, the temporal resolution of MWL tracking decreases as it takes time to estimate MWL. Therefore, the effect of the time window duration should be investigated to maximize the accuracy of MWL estimation. The aim of this study is to investigate the appropriate time window range for calculating eye movement parameters in mental workload (MWL) estimation during automobile driving. Furthermore, the accuracy of MWL estimation and correlation between the subjective and estimated MWL are compared for different time windows using experimental data obtained in the previous study [2].

2 Methods

2.1 Experimental Conditions

Twelve Japanese students with an average age of 21.6 ± 0.51 years and having drivers' licenses participated in this experiment. They performed driving tasks in an urban city course on a driving simulator (UCwin/Road Ver.13 Driving Sim, FORUM8 Inc.) (see Fig. 1); their eye and head movements were simultaneously measured with an image sensor (B5T-007001, Omron Inc.) by controlling their MWL through the N-back task [3]. In the N-back task, the participants were required to keep answering aloud the N-th previous digit in a sequence of digits that was read out consecutively. The N-back task had five difficulty levels: none, 0-back, 1-back, 2-back, and 3-back. The subjective MWL was measured using the national aeronautics and space administration task load index (NASA-TLX) [4, 5]. The adaptive weighted workload (AWWL) score [5] was used as the total score. The AWWL score is calculated as the weighted sum of the six scales with the weights of 6, 5, 4, 3, 2, and 1 in a decreasing order of the scales. The higher the value, the higher is the subjective MWL.

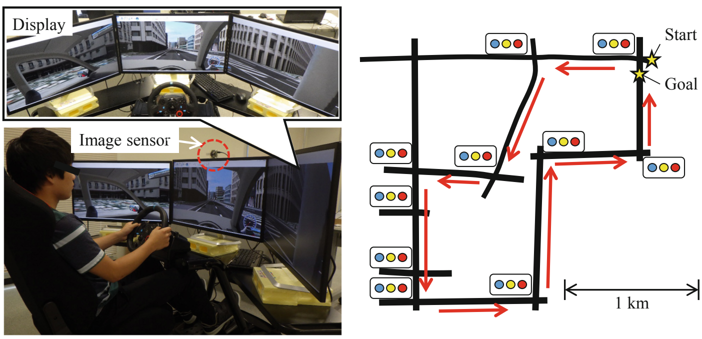


Fig. 1. Driving simulator and driving route.

2.2 Quantification of MWL by One-Class Support Vector Machine

The gaze angle, head angle, and degree of eye closure during the driving task were measured using an image sensor. Thereafter, the four eye movement parameters (i.e., SD of gaze angle, SD of eyeball rotation angle, sharing rate of head movement, and blink frequency) were calculated by changing the time window to 30–150 s in increments of 30 s (Fig. 2). Anomaly detectors for MWL were constructed using the one-class support vector machine (OCSVM) for each participant and each increment of the time window. The OCSVM creates a decision function that takes a non-negative value in the area containing a large volume of training data and a negative value in the other areas. We used the OCSVM implemented in scikit-learn 0.23.2. The radial basis function (RBF) kernel was used as the kernel function, and the coefficient of the kernel γ was set as $\gamma = 0.25$. In addition, the upper bound on the fraction of training errors ν was set as $\nu = 0.01$. The two hyper parameters of the OCSVM were heuristically determined.

In total, 50% of the "none" data were randomly used as the training data for the OCSVM, and the remaining "none" and "0-back" to "3-back" data were used as test data. Note that the training and the test data were normalized based on the means and SDs of the four eye movement parameters in the training dataset.

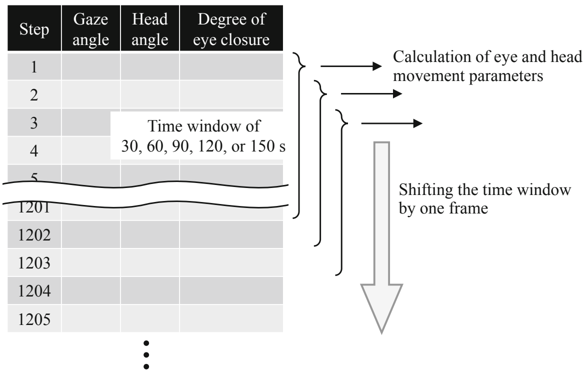


Fig. 2. Calculation of eye and head movement parameters.

2.3 Analysis

For each window length condition, we calculated the area under the curve (AUC) for the binary classification between "none" and "3-back" data, the percentage of anomalous data, and the distance of each eye movement parameter data from the decision boundary. A one-way analysis of variance (ANOVA) was conducted to investigate the effects of the time window on the AUC; Tukey's post-hoc tests were carried out to compare the levels of the time window.

3 Results

Figure 3 shows the average AUCs for each time window. The ANOVA revealed that the effect of time window was significantly associated with the AUC at 1% significance level. The time window of 30 s had significantly lower AUC compared with other time windows.

The ratios of the anomalous data for each time window are shown in Fig. 4. The ratio of the anomalous data monotonically increased with the increase in the task difficulty in the range from "0-back" to "3-back." In addition, the anomaly ratio also increased with the increase in the duration of the time window with respect to the same task difficulty except "None." Especially for the time window duration of 150 s, the ratio of the anomalous data became almost 100% for "1-back" to "3-back" tasks.

Figure 5 shows the correlation coefficients between the subjective MWL (i.e., AWWL score of NASA-TLX) and the distance of each eye movement parameter data from the decision boundary of OCSVM ($N = 5$ task difficulties $\times$ 12 participants = 60). The correlation coefficients were significant at 1% significance level irrespective of the time window. The correlation coefficient monotonically increased between 30 s and 120 s and decreased at 150 s. Figure 6 shows the relationship between AWWL scores and the distance of each eye movement parameter data from the decision boundary at 120 s time window, which has the highest correlation coefficient.

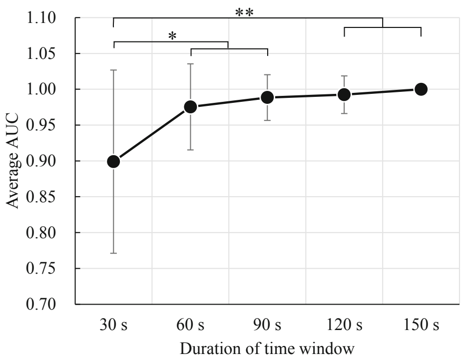
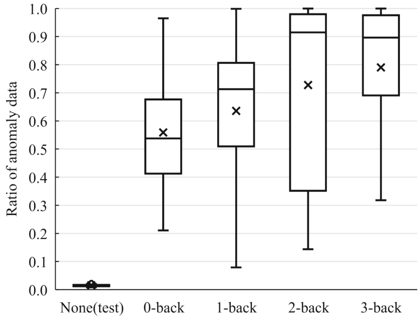


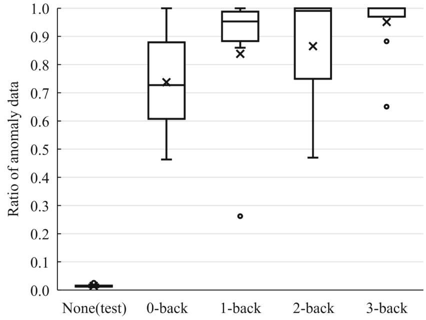
Fig. 3. Relationship between the time window and average AUC. Error bars represent the standard deviations. This graph includes the results of the post-hoc test; * and ** represent $p < 0.05$ and $p < 0.01$, respectively.

4 Discussion

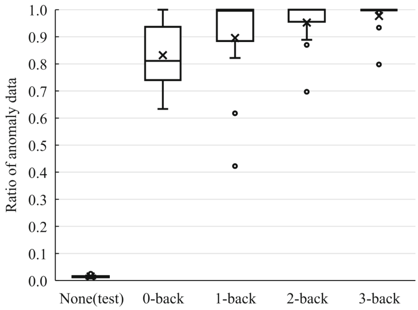
As shown in Fig. 4, the ratio of the anomalous data monotonically increased with the task difficulty irrespective of the time window; therefore, the time window of 30–150 s can quantify the MWL during driving. The participants showed different abilities for the N-back task; thus, MWL from the same N-back task is different for each participant. Therefore, the ratio of the anomalous data is expected to have some variability among



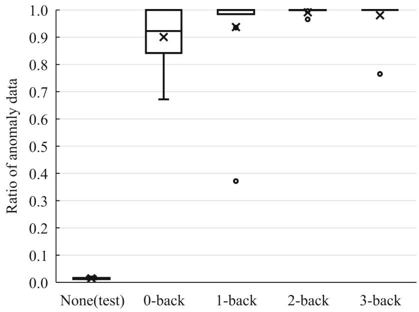
(a) 30 s



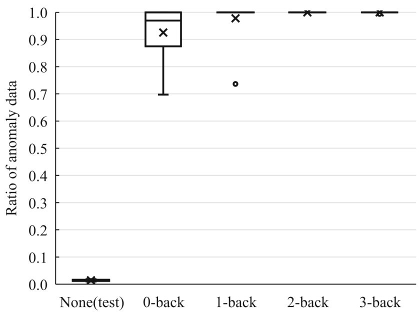
(b) 60 s



(c) 90 s



(d) 120 s



(e) 150 s

Fig. 4. Box plots of ratio of anomaly data for each time window. The bottom and top edges of the box represent the first and third quartiles (Q1 and Q3), respectively, and the band in the box is the second quartile (Q2) or the median. The white dot represents an outlier; the threshold for outlier determination was less than Q1 – 1.5 × IQR (IQR = interquartile range) and higher than Q3 + 1.5 × IQR. The bottom and top of the whisker represent the maximum and minimum excluding the outliers, and the cross represents the average value.

the participants, especially in the relatively easy N-back task. By contrast, the “3-back” task was the most difficult one, resulting in a considerably high MWL for almost all the

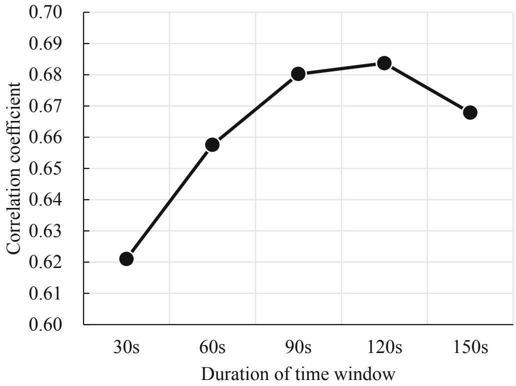


Fig. 5. Correlation coefficient between the normalized AWWL score and the distance from the decision boundary for each time window.

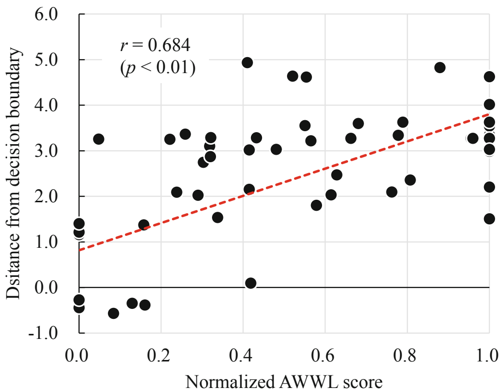


Fig. 6. Relationship between the normalized AWWL score and the distance from the decision boundary for the time window of 120 s.

participants. However, the average of the ratio of the anomalous data for the "3-back" was approximately 80% when the time window was 30 s (Fig. 4(a)). In addition, the detection ability for the time window of 30 s is lower compared with other time windows. This is because 30 s is relatively short for calculating the eye movement parameter, and the detection accuracy decreases due to the effect of the noise. Therefore, the time window should be more than 30 s.

The time window 120–150 s has approximately 100% anomalous data ratios, not only for the "3-back" but also for the "1-back" and "2-back" tasks. It would be more natural for the ratio of the anomalous data to increase gradually with the N-back task because the MWL from the relatively low-difficulty task may have variability across the participants. In addition, too long a window length (i.e., 150 s) impairs the correlation between the objective and subjective MWL evaluation. The time window of 150 s masks the difference for "1-back" to "3-back" tasks, whereas the anomaly ratio should increase gradually. Therefore, 150 s is slightly too long as the time window.

Considering the aforementioned points, we concluded that 60 s to 120 s is an appropriate time window duration for MWL evaluation.

5 Conclusions

We found that the range of 60 s to 120 s is the appropriate time window duration for calculating eye movement parameters in mental workload (MWL) estimation during automobile driving.

References

1. Statistics about Road Traffic/Traffic accidents situation. https://www.e-stat.go.jp/en/stat-search/files?page=1&layout=datalist&toukei=00130002&tstat=000001027458&cycle=7&year=20190&month=0&stat_infid=000031910507&tclass1val=0. Accessed 21 Jan 2021
2. Chihara, T., Kobayashi, F., Sakamoto, J.: Evaluation of mental workload during automobile driving using one-class support vector machine with eye movement data. Appl. Ergon. **89**, 103201 (2020)
3. Ross, V., Jongen, E.M.M., Wang, W., Brijs, T., Brijs, K., Ruiter, R.A.C., Wets, G.: Investigating the influence of working memory capacity when driving behavior is combined with cognitive load: an LCT study of young novice drivers. Accid. Anal. Prevention **62**, 377–387 (2014)
4. Hart, S.G., Staveland, L.E.: Development of NASA-TLX (Task Load Index): results of empirical and theoretical research. Adv. Psychol. **52**(C), 139–183 (1988)
5. Miyake, S., Kumashiro, M.: Subjective mental workload assessment technique – an introduction to NASA-TLX and SWAT and a proposal of simple scoring methods. Jpn. J. Ergon. **29**(6), 399–408 (1993). (in Japanese)