Artificial Intelligence In Medicine 91 (2018) 39–48

Contents lists available at ScienceDirect

Artificial Intelligence In Medicine

journal homepage: www.elsevier.com/locate/artmed

Detecting mental fatigue from eye-tracking data gathered while watching video: Evaluation in younger and older adults

Yasunori Yamada*, Masatomo Kobayashi

IBM Research – Tokyo, 19-21, Nihonbashi, Hakozaki-cho, Chuo-ku, Tokyo 103-8510, Japan

A R T I C L E I N F O

Keywords:
Mental fatigue
Stress
Eye movement
Free viewing
Visual attention model
Elderly health

A B S T R A C T

Health monitoring technology in everyday situations is expected to improve quality of life and support aging populations. Mental fatigue among health indicators of individuals has become important due to its association with cognitive performance and health outcomes, especially in older adults. Previous models using eye-tracking measures allow inference of fatigue during cognitive tasks, such as driving, but they require us to engage in specific cognitive tasks. In addition, previous models were mainly tested by user groups that did not include older adults, although age-related changes in eye-tracking measures have been reported especially in older adults. Here, we propose a model to detect mental fatigue of younger and older adults in natural viewing situations. Our model includes two unique aspects: (i) novel feature sets to better capture fatigue in natural-viewing situations and (ii) an automated feature selection method to select a feature subset enabling the model to be robust to the target's age. To test our model, we collected eye-tracking data from younger and older adults as they watched video clips before and after performing cognitive tasks. Our model improved detection accuracy by up to 13.9% compared with a model based on the previous studies, achieving 91.0% accuracy (chance 50%).

1. Introduction

Health monitoring technology in a smart environment such as adaptive workplaces and smart houses has been increasingly recognized as a way of improving health outcomes as well as cognitive and behavioral performance [1]. Especially considering the growing demand for health monitoring for older adults due to a rapidly aging population, monitoring technology is expected to support *aging in place* and help enable individuals to manage their own health, with the dual aim of increasing quality of life and reducing healthcare costs [2]. Previous studies on health monitoring have suggested behavioral features in speech, motion, and gaze as a means of inferring an individual's state, such as physical conditions [3], stress [4,5] and mental workload [6], measuring behavioral characteristics, such as sleep quality [7] and social activities [8], and screening for neurodegenerative diseases, such as dementia [9–12], depression [13], and Parkinson's disease [14]. These monitoring technologies can keep track of daily changes in health and detect early signs of disease. Being capable of inferring unutilized information related to an individual's health holds promise for providing better health care and enhancing well-being.

One aspect of an individual's daily health status that has yet to be utilized is mental fatigue, which refers to the feeling people might experience during or after cognitive activities [15]. Mental fatigue is becoming an increasingly serious health and social problem, and it comes at a huge public health cost [16]. In the workplace, mental fatigue is known to affect cognitive and behavioral performance [17]. In fact, mental fatigue has been suggested as one of the most frequent causes of accidents and errors in the workplace [18]. Recent analyses have reported that the cost of fatigue-related accidents and errors in the US may reach as a high as $31.1 billion [19,20]. From the perspective of an individual's health, mental fatigue is a warning sign of harmful accumulations of stress that can have a detrimental effect on one's health [21]. Furthermore, in the context of health of the elderly, it has garnered increasing attention as recent longitudinal studies have shown an association of mental fatigue with cognitive decline and daily functional deficits in later life [22,23].

Previous studies on monitoring mental fatigue by using unobtrusive methods have primarily focused on using eye-tracking measures during cognitive tasks such as driving [18,24,25]. Some studies investigated how eye-tracking measures change with the duration of the cognitive task to identify sensitive measures indicating an increase in mental fatigue [26–28]. Others built models by combining these eye-tracking measures and succeeded in detecting mental fatigue during a specific cognitive task [29,30]. However, no study has yet investigated the association between mental fatigue from eye-tracking measures in natural-viewing situations when the individual is not performing cognitive

* Corresponding author.
E-mail address: ysnr@jp.ibm.com (Y. Yamada).

https://doi.org/10.1016/j.artmed.2018.06.005
Received 8 November 2017; Received in revised form 11 May 2018; Accepted 26 June 2018

tasks. A model capable of detecting mental fatigue from eye-tracking data in natural-viewing situations would extend the scope of application to monitoring fatigue in various everyday situations. Moreover, it could be used to infer fatigue induced by not only specific cognitive visual tasks, but also various factors such as cognitive auditory tasks, multiple cognitive tasks, or poor health [18].

In addition, recent studies have pointed out the need to test whether previous models for inferring mental fatigue using eye-tracking measures can be applied to aging populations because age-related changes in eye-tracking measures have been reported [29,31,32]. As mentioned above, mental fatigue has become important for monitoring the health of older adults [22,23]; thus, a model that enables us to monitor mental fatigue in a wider age group including older adults would be very useful.

In this paper, we propose a novel model to detect mental fatigue of younger and older adults in natural viewing situations. Our model's uniqueness lies in (i) novel feature sets to better capture mental fatigue in natural viewing situations and (ii) an automated feature selection method to select a feature subset enabling the model to be robust to the target's age. We collected eye-tracking data from 18 participants while they watched video clips and tested whether our model could detect mental fatigue induced by auditory cognitive tasks. With eye-tracking data of individuals watching only 30 s worth of video, our model could determine whether that person was fatigued or not with 91.0% accuracy in ten-fold cross-validation (chance 50%). To make a comparison with a model based on the existing work, we also built a model with only feature sets used in a previous study, where the detection accuracy was 77.1%. We also collected eye-tracking data from 11 additional participants while they watched video clips without engaging in cognitive tasks, and confirmed that our model captured changes resulting from mental fatigue induced by the cognitive tasks, not just sequence effects of repetitive video watching. A preliminary version of this paper appeared as [33].

Our contributions are as follows:

- We developed a model to detect mental fatigue of younger and older adults in natural viewing situations.
- We collected eye-tracking data from younger and older adults who watched video clips before and after performing cognitive tasks.
- Through a comparison with a model based on previous studies, we demonstrated that our model could robustly detect mental fatigue induced by cognitive tasks, despite there being age-related changes in the eye-tracking measures.

2. Related work

In this section, to determine the sort of eye-tracking measures that we should focus on, we describe the existing fatigue detection technologies and measures correlated with mental fatigue. We also explain recent studies on the association with eye-tracking data in natural viewing situations and an individual's health state.

2.1. Fatigue detection technology

Many attempts at developing fatigue detection technology have focused on either determining fatigue levels by using specific test tasks, i.e., fitness-for-duty tests [34,29], or monitoring correlates of fatigue during cognitive tasks [30,25]. The former sort of approach uses neurobehavioral tasks for assessing executive functions such as vigilance or hand-eye coordination; their test tasks typically take up to 10 min and have no practice/learning effects [29]. For example, these test tasks are performed before and/or in the middle of driving in order to evaluate fatigue [29]. Although this enables one to estimate an individual's mental fatigue regardless of what kind of cause induced the mental fatigue, it requires the person to perform the test task every time the mental fatigue evaluation is to be made. On the other hand, the latter sort of approach aims to infer mental fatigue during cognitive tasks without any additional test tasks. Most of the studies on this approach built models for inferring mental fatigue by using task performance and/or psychophysiological measures such as eye-tracking measures that are known to be correlated with fatigue during the targeted cognitive task [25,19]. These studies have achieved high detection accuracy, and several commercial products based on them are already available [29,30]. However, because their systems target fatigue during specific cognitive tasks, we need to examine the question of whether they can be used to estimate mental fatigue during other cognitive tasks.

Both approaches typically use task performance and/or psychophysiological measures as fatigue-correlated measures. The measures related to task performance are relatively easy to obtain and interpret. However, their validity remains controversial because recent studies suggest that mental fatigue does not always impair task performance; specifically, they reported that people can keep performing prolonged tasks even when multiple other measures including eye-tracking measures show accumulating mental fatigue [35]. In addition, if one uses task performance as criteria for inferring mental fatigue, we need to determine how to calculate the performance measures and build a model for each cognitive task. Thus, recent studies have aimed at identifying psychophysiological measures correlated with mental fatigue and developing models by combining them in various ways [25,30].

2.2. Fatigue correlated psychophysiological measures

The most advanced psychophysiological measures are electroencephalography (EEG) and eye-tracking measures. EEG measures that capture changes in brain waves have been shown to be valid biomarkers of mental fatigue [29,36]. Specifically, multiple studies have shown that as an individual grows fatigued, slow wave activity such as theta and alpha activity increases over the entire cortex [29,36]. Although EEG measures have high accuracy for offline fatigue monitoring, they require the time-consuming application of obtrusive EEG sensors to the head of the individual, and thus, it would be difficult to use them for health monitoring applications in everyday situations, at least for now. On the other hand, eye-tracking measures have an advantage in their unobtrusiveness, and they were used by most of the previous studies and applications aiming to monitor mental fatigue [29]. The eye-tracking measures for inferring mental fatigue typically include indices associated with pupil measures, blinking, and oculomotor-based metrics [28,29,18].

The relationship between pupil measures and mental fatigue has been studied since the 1930s [37], and multiple fatigue correlated measures have been reported, for example, pupil diameter, and constriction velocity and amplitude [29,26,18]. Blinking is one of the visual behaviors that can be easily observed when an individual is fatigued. Several studies have shown that increased fatigue results in longer and more frequent blinking [27]. Other studies reported other measures related to increased mental fatigue, such as blink velocity and interval [25,24]. Schleicher and colleagues experimented in a simulated traffic situation and found that blinking was the best indicator of fatigue compared with other eye-tracking measures including oculomotor-based metrics such as saccadic parameters and fixation durations [24]. Among the various oculomotor-based metrics, saccadic metrics have received the most attention as a means of inferring mental fatigue [28,38]. Multiple studies have investigated the validity and sensitivity of saccadic metrics, including saccade velocity, duration, and amplitude, as indexes of an individual's fatigue [25,28]. Di Stasi and colleagues demonstrated that saccadic eye movement parameters, particularly the peak velocity, are sensitive indicators of mental fatigue in experiments using different tasks, including simulated air traffic control, simulated driving, and simulated laparoscopic surgery [25,28]. In addition to the saccadic metrics, recent advances in eye tracking technologies have enabled us to record eye positions at high speed and to extract from data drifts and microsaccades (small-magnitude saccades produced during attempted fixation) [39,19]. These studies have shown that microsaccade velocities decrease as fatigue increases in a range of tasks in different settings [39,19].

The pupil measures and blink data can be obtained from images by using computer vision techniques [30]. Regarding oculomotor-based metrics, while low-cost eye trackers are now becoming available, a webcam-based tracking system has also been proposed [40]. This webcam-based system enables eye-tracking data including oculomotor-based metrics related to saccades and fixations to be collected even in a crowdsourcing setting, and the data has a quality comparable to data gathered by using low-cost eye-trackers in traditional lab settings [40]. From this perspective, oculomotor-based metrics related to saccades and fixations could also be used to develop a model for inferring mental fatigue in everyday situations. In contrast, metrics related to drifts and microsaccades cannot be measured using low-cost eye trackers and require specialized expensive hardware with high sampling rates such as 1000 and 2000 Hz, making it difficult to use in a living environment. Thus, for the time being, drift and microsaccade measurements are still not applicable.

These eye-tracking measures have been experimentally tested mainly for people below 50 years of age. In contrast, many studies have reported age-related changes in eye-tracking measures including measures used for inferring mental fatigue in previous studies such as pupil diameter and responses, eye movement patterns, and saccades [29,32,41]. For example, large-scale studies on age-related changes in saccade dynamics reported that saccadic parameters including latency and velocity are relatively stable throughout the middle years up to age 50, but then change in later years [31]. However, no study has investigated the question of whether eye-tracking measures can be used to infer mental fatigue in older adults [29]. To develop a model that is robust to age-related changes in eye-tracking measures, we need to test a model in individuals including older adults.

2.3. Eye-tracking measurements in natural viewing situations

The previous studies and applications for inferring mental fatigue mainly focused on correlates of fatigue when a person engages a specific cognitive task, and no study has yet developed a model that allows us to infer mental fatigue in natural-viewing situations when the target is not performing cognitive tasks.

Instead, recent studies have attempted to screen for brain diseases by using eye-tracking data in natural viewing situations [42,38,10,43]. Crabb and colleagues demonstrated that patients with neurodegenerative eye disease can be separated from healthy controls by using eye-tracking data collected while people freely watched TV-type films [38]. Similarly, Tseng and colleagues devised a three-class classification model that could discriminate between children with attention deficit hyperactivity disorder, fetal alcohol spectrum disorder, and controls in natural viewing situations where participants simply watched video clips (simulating the situation of watching a TV program) [42]. In these studies, the eye tracking data were characterized by the distribution of eye movements and a saliency model in addition to basic oculomotor-based metrics such as saccade velocity and fixation duration.

The saliency model was originally proposed as a biologically inspired computational model related to human attention based on feature integration theory [44,45]. It has been widely used for analyzing eye-tracking data collected in natural viewing conditions [46,47]. This model attempts to explain how human eye movements are guided by bottom-up attention—in other words, stimulus-driven or reflexive eye movements [45]. The term bottom-up attention is used in contradistinction to top-down attention, which refers to the voluntary allocation of attention [45,48]. The saliency model tries to quantify the contribution of bottom-up attention to eye movement based on the prediction performance of fixations calculated by using only visual stimuli [46,47]. Although there has been no investigation on the associations between mental fatigue and saliency-based eye movement features, it seems likely that reflexive eye movement guided by bottom-up attention increases with mental fatigue. In fact, psychological studies have shown that fatigued participants had difficulty sustaining attention and ignoring irrelevant information [15], supporting the hypothesis that eye movement might be more affected by bottom-up attention when an individual is fatigued.

3. Mental fatigue detection model

On the basis of the related work described in the previous section, we decided to focus on six types of eye-tracking measures that would be useful for inferring mental fatigue: feature sets related to pupil measures, blinking, oculomotor-based metrics, gaze allocation, eye movement directions, and a saliency model. The first three feature sets were used in previous studies on mental fatigue during cognitive tasks [24,25,18]. The other three feature sets have been used for characterizing eye movements in natural viewing situations as well as inferring brain diseases [42,38,10], although they have not been used before for detecting mental fatigue.

Furthermore, we investigated age-related changes in eye-tracking measures that have been used for inferring mental fatigue in previous studies. To build a model capable detecting mental fatigue of individuals including younger and older adults, we need to identify eye-tracking features that enable the model to be robust to age-related changes through a feature selection method.

Our model is summarized in Fig. 1. In this study we treated a binary

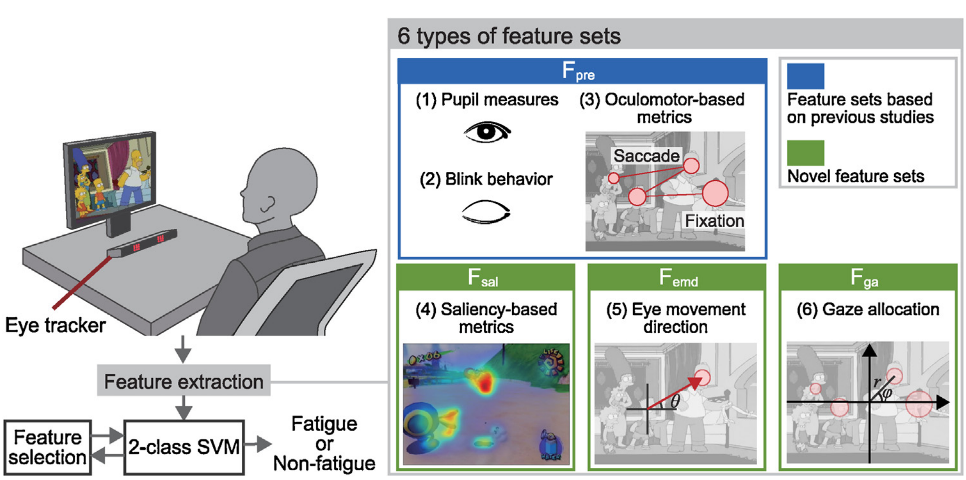

Fig. 1. Overview of our fatigue-detection model. Our model first extracts six types of feature sets from eye-tracking data collected while participants watch video. Using a subset of the features selected by a feature selection method, a two-class classifier using support vector machine (SVM) model estimates whether that person is fatigued or not.

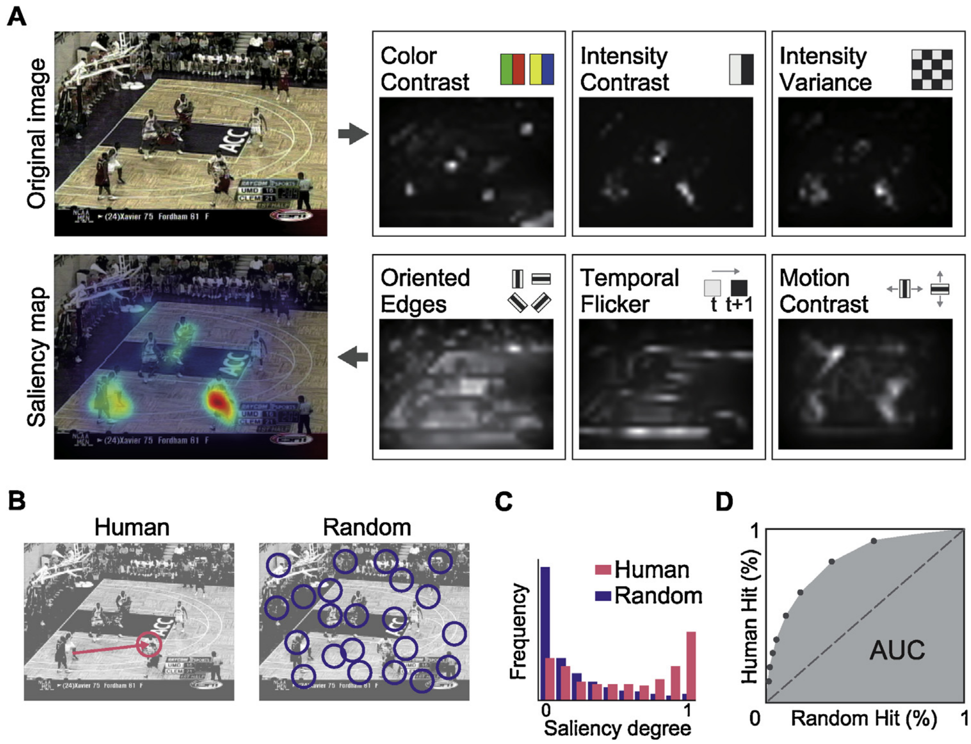

Fig. 2. Workflow of how saliency-based features are extracted. (A) The saliency model first generates conspicuity maps in terms of six different low-level features, and these maps are linearly combined and normalized to form a saliency map. The saliency map represents a topographic map of conspicuity for every location in each video frame, highlighting locations that may attract attention in a stimulus-driven manner. (B) For each saccade endpoint, a saliency map value is sampled. At the same time, map values are randomly sampled from the same saliency map. (C and D) Using these values with all the saccades, histograms are generated from saliency values of human and random fixations and summarized using ordinal dominance analysis for calculating the AUC score.

state of mental fatigue, and proposed a model to capable of inferring mental fatigue. To build our model, we first extracted 181 quantitative features, categorized into six feature sets that may change according to an individual's state of mental fatigue. We next used a two-class classifier for inferring mental fatigue by using a subset of the features selected by a feature selection method through recursive evaluation and selection to avoid over-fitting.

For a two-class classification model for detecting mental fatigue, we used support vector machine (SVM) models [49,50] with a radial basis function kernel as follows: $K(\boldsymbol{x}_i, \boldsymbol{x}_j) = \exp(-\gamma||\boldsymbol{x}_i - \boldsymbol{x}_j||^2)$. We set $\gamma = b_{\mathrm{SVM}}/n_{\mathrm{f}}$, where n_{f} is the number of features and b_{SVM} is a hyper-parameter. We used the algorithm for SVM implemented in MATLAB (MathWorks Inc., Natick, MA) and LIBSVM toolbox [50].

To avoid overfitting of the model and identify features that enable the model to be robust to age-related changes, we performed the feature selection through recursive evaluation and selection. One of the well-known algorithms for this is support vector machine recursive feature elimination (SVM-RFE) in the wrapper approach [51]. SVM-RFE is a powerful feature selection algorithm, and it has been used with great success in pattern recognition applications. It uses criteria derived from the coefficients in SVM models to assess features and iteratively discards the weakest features until all of them have been eliminated. However, when the candidate feature set has highly correlated features, the ranking criterion of SVM-RFE tends to be biased, which would have a negative effect on the results. In fact, our feature set extracted from eye tracking data contained highly correlated features such as features related to saccade duration, amplitude, and velocity. We thus used an improved SVM-RFE algorithm with a correlation bias reduction strategy in the feature elimination procedure [52].

3.1. Feature extraction

The oculomotor-based features consisted of nine features: saccade amplitude, saccade duration, saccade rate, inter-saccade interval (mean, standard deviation, and coefficient of variance), saccadic mean velocity (mean and median), and fixation duration. The previous studies showed that saccadic mean and peak velocity decreased with increasing mental fatigue [25,28]. However, low-cost eye trackers do not reliably detect saccadic peak velocity because of their relatively low sampling rate. Accordingly, we used only saccadic mean velocity, so that our model could be implemented using low-cost eye trackers such as webcam-based tracking systems [40,53].

We calculated seven features related to blinking: blink duration, blink rate, blink duration per minute (the total time of all durations), and inter-blink interval (mean, standard deviation, and coefficient of variance).

The pupil measures were subdivided into six features related to pupil diameter, constriction velocity, and amplitude of each eye, and nine features related to the coordination of the pupil diameters of both eyes. Of these nine features, one was computed using Pearson's correlation coefficient. The other eight features were extracted using the phase locking value [54], which can identify transient synchrony over shorter time scales than Pearson's correlations can. We used the mean and maximum values of the phase locking values with four different time windows (5, 10, 30, and 60 frames).

The fourth feature set was calculated from a time-series of gaze allocation. We first converted gaze allocation values into radius and angle (r, ϕ) values in a polar coordinate system situated at the center of the display. We then defined time series of gaze allocations during all periods as $(\boldsymbol{r}, \boldsymbol{\phi})_{\mathrm{all}}$ and those only during fixation periods as $(\boldsymbol{r}, \boldsymbol{\phi})_{\mathrm{fx}}$. We discretized each time series with k bins of uniform width. We set $k = 8$ for $\boldsymbol{r}_{\mathrm{all}}$ and $\boldsymbol{r}_{\mathrm{fx}}$, $k = 36$ for $\boldsymbol{\phi}_{\mathrm{all}}$, and $k = 12$ for $\boldsymbol{\phi}_{\mathrm{fx}}$. As features, we used the probability of each bin and entropy estimated using these histograms and also calculated the mean and median values of $\boldsymbol{r}_{\mathrm{all}}$ and $\boldsymbol{r}_{\mathrm{fx}}$. In total, we obtained seventy-two features from the gaze-allocation data.

The fifth feature set related to eye-movement directions was calculated in a similar manner to the gaze allocation features. We discretized the time series of eye-movement directions θ during all periods and during saccades periods into 12 and 36 bins of uniform width, respectively. We then computed the probability of each bin and entropy estimated using these histograms as features. In total, we obtained fifty features.

The sixth and final set consisted of features extracted using a saliency model (Fig. 2). In particular, we used the graph-based visual

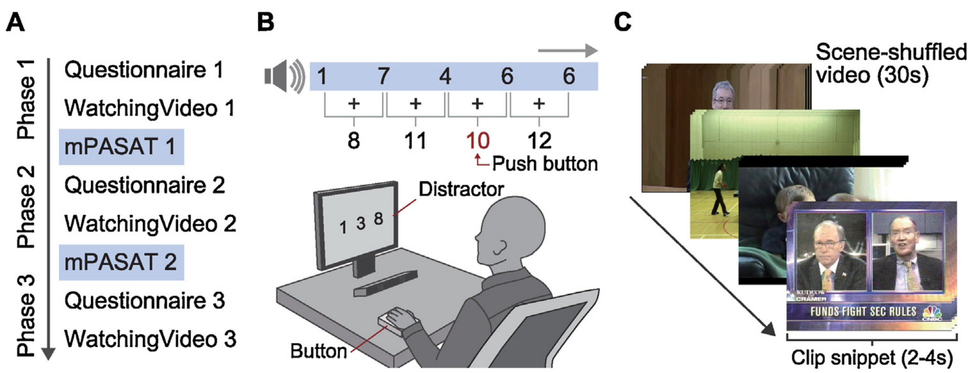

Fig. 3. Experimental setup: (A) overall procedure, (B) mental calculation task called mPASAT, and (C) examples of scene-shuffled video clips.

saliency model proposed by Harel et al. [55], because it has been shown to have high prediction power for fixations on several datasets and can be applied to dynamic scenes (videos). This model first generates conspicuity maps in terms of six low-level features: intensity contrast, color contrast, intensity variance, oriented edges, temporal flicker, and motion contrast. Next, conspicuity maps of the six different low-level features are linearly combined and normalized to form a saliency map. The saliency map represents a topographic map of conspicuity for every location in each video frame, highlighting locations that may attract attention in a stimulus-driven manner.

We then compute the prediction performance of the saliency map by using the area under the curve (AUC) of the receiver operating characteristic, as follows. First, we obtained a saliency map value at the saccade endpoint when a participant started a saccade, and we randomly sampled 100 map values from the same saliency map. These values were normalized to values from 0 to 1 relative to the minimum and maximum values in the saliency map. Using these values with all the saccades, human fixations were considered to be the positive sets and random ones the negative sets. The saliency map was then treated as a binary classifier to separate the positive samples from the negatives. By incrementing a threshold from 0 to 1, we plotted the true positive (human hits) rate versus the false positive (random hits) rate and obtained the ROC curve. Therefore, the AUC represents the predictive accuracy of the saliency model based on the endpoints of the saccades of the participants. For example, an AUC of 0.5 indicates chance level and 1.0 means perfect prediction.

We used the AUC score and the probability of low-/medium-/high saliency map values at the saccade endpoints as saliency-based features. The saliency-based features can be calculated using not only the saliency map, but also each conspicuity map of the six low-level features. Accordingly, we obtained $4 \times 7 = 28$ saliency-based features in total. For more details, we refer the reader to the original papers [42,55,46].

4. Experiments

We conducted two experiments. In the first experiment, we collected eye-tracking data from younger and older adults in a situation in which they watched video clips (simulating the situation of watching a TV program) before and after performing auditory cognitive tasks. We then investigated whether our model could detect mental fatigue from the eye-tracking data. In the second experiment, we aimed to confirm whether our model could capture the difference in eye-tracking data resulting from increased mental fatigue induced by the cognitive tasks, not simply from the sequence effects of repetitive video watching. To this end, we collected eye-tracking data from other participants who did not engage in the cognitive tasks. Specifically, they were asked to watch the video clips in the same order as the first experiment, and rest between phases without engaging in cognitive tasks.

4.1. Participants

For the first experiment, we collected data from 20 participants (eight females, 12 males; 24–76 years; mean ± SD age, 47.5 ± 20.5 years). Among the participants, nine (five females, four males) were more than 50 years of age, and they were recruited from local senior communities. For the second experiment, we collected data from 11 other participants (two females, nine males; 23–57 years; mean ± SD age, 29.7 ± 9.8 years). Eye tracking data from two participants for the first experiment (one female and male; the female was more than 50 years old) were excluded from our analysis because of problems calibrating the eye tracker. Thus, our sample size was $N = 18$ and $N = 11$ for the first and second experiments, respectively. All participants were well-rested and in good health, as measured by self-reports, and they had normal or corrected-to-normal vision. They were unaware of the purpose of the experiments.

4.2. Experimental design and procedure

The experimental procedure for the first experiment is summarized in Fig. 3. Participants of the first experiment performed a 17-minute mental calculation task designed to induce mental fatigue two times. They were asked to take questionnaires and watch video clips prior to and following each mental calculation task (Fig. 3A). Prior to the experiment, all participants were given oral instructions about the experiments and allowed to practice the mental calculation task. Participants of the second experiment followed the same procedure as the first experiment with the exception that they rested instead of performing the mental calculation task.

We used numerical rating scales to measure the current (“right now, at this moment”) perceived intensity of feelings about mental and physical fatigue, sleepiness, and motivation. The intensity was scaled from zero to ten, with zero indicating an absence of those feelings and ten indicating the strongest feeling ever experienced. Numerical rating scales were used because they are considered superior for the assessment of unidimensional subjective feelings compared with other methods such as visual analogue scales [56].

The participants were asked to watch video clips approximately 5 min in length during each phase of the experiment. Stimuli were shown on a 20-inch computer monitor and presented at 30 Hz with a resolution of 1600×1200 pixels. Participants were seated approximately 80 cm away from the monitor so that scenes subtended approximately $28^{\circ} \times 22^{\circ}$ of the usable field-of-view. As in previous studies [42], they were instructed to simply “watch and enjoy the videos.”

As an auditory cognitive task to induce mental fatigue, we used a modified version of the paced auditory serial attention test (mPASAT) [57] (Fig. 3B). The mPASAT requires several cognitive functions such as working memory, attention, and arithmetic capabilities while imposing a high cognitive workload. Several studies have reported that the mPASAT induces mental fatigue in participants [57]. In this task,

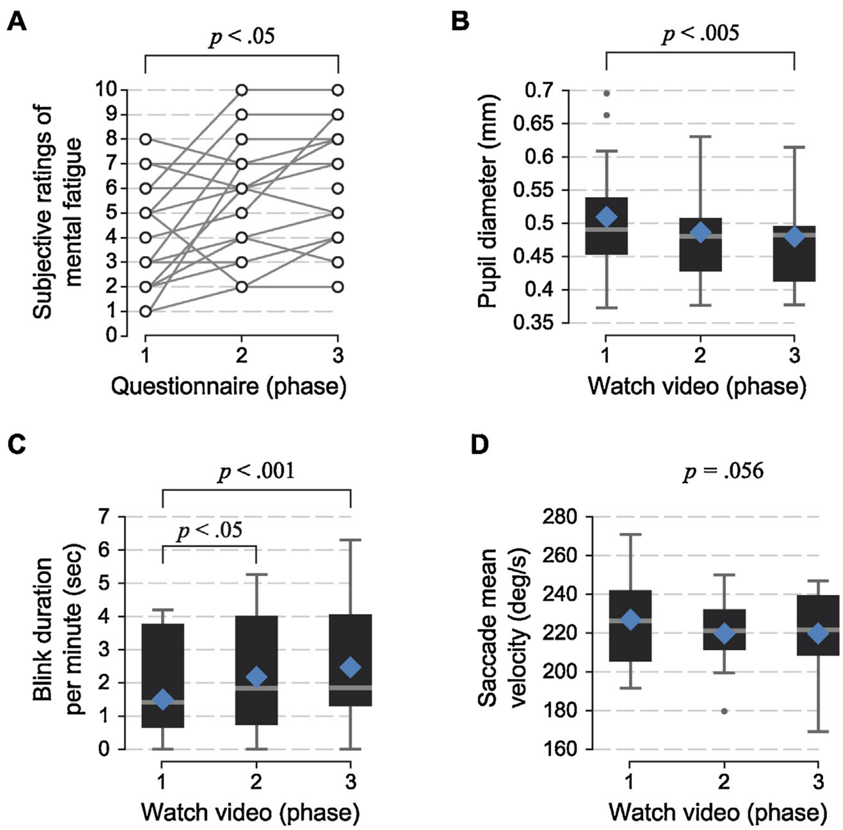

Fig. 4. Changes in subjective and objective measures of mental fatigue after performing mPASAT. (A) Subjective ratings of mental fatigue on an 11-point numerical rating scale from 0 to 10. We found a significant increase in mental fatigue in subjective ratings from phase 1 to 3 ($p < .05$, one-way repeated-measures Friedman non-parametric ANOVA with Dunn's post-hoc test). (B, C, and D) Right pupil diameter, blink duration per minute, and saccade mean velocity. These are widely used as fatigue-correlated measures. Right pupil diameters significantly decreased from phase 1 to 3 ($p < .005$,), and blink duration per minute significantly increased from phase 1 to 2 and 3 (from phase 1 to 2, $p < .05$; from phase 1 to 3, $p < .001$). Saccade mean velocity slightly decreased over the phases, but there was no significant difference ($p = .056$). For statistical analyses, we performed one-way repeated measures ANOVA and post hoc Bonferroni multiple comparisons. Boxes denote the 25th (Q1) and 75th (Q3) percentiles. The line within the box denotes the 50th percentile, while whiskers denote the upper and lower adjacent values that are the most extreme values within Q3 + 1.5(Q3 − Q1) and Q1 − 1.5(Q3 − Q1), respectively. Filled circles show outliers, and squares represent mean values.

participants listened to a series of numbers ranging from one to nine. They were asked to add the number they had just heard to the number they had heard before and then to press a button whenever the sum of the two consecutive numbers equaled ten. One phase consisted of five three-minute on-periods and four 30-second off-periods for a total of 17 min. Each number was presented for 1.5 s; 120 numbers were thus presented during each three-minute on-period. Participants were also asked to visually focus on three numbers on the display, which randomly changed every 0.5 s. These visual numbers were intended to distract and interfere with the primary auditory task, thereby increasing the complexity and attentional demands of the task in order to induce further mental fatigue. To summarize: participants were instructed to listen to a series of numbers presented via the speaker and perform the serial addition task while simultaneously keeping their eyes open and focusing on the numbers displayed on the screen.

4.3. Eye-tracking data acquisition and stimuli

The participants' eye movements and pupil data were recorded using a noninvasive infrared EMR ACTUS eye tracking device at a sample rate of 60 Hz (nac Image Technology Inc.; spatial resolution for eye movements and pupil diameter less than 0.5° and 0.1 mm, respectively). The eye tracker was calibrated using 9-point calibration at the beginning of each recording phase.

To simulate the situation of watching a TV program, we used video clips made in the same manner as previous studies that investigated how brain diseases affect eye movements in natural viewing situations [48,42] (Fig. 3C).

Each five-minute phase consisted of nine scene-shuffled videos (SVs), approximately 30 s each. Between the SVs, there were five-second off-periods for rest. The SVs were made by assembling randomly extracted snippets from video clips. The lengths of the snippets were determined so that they were within the range of typical television programs [58,59]. For example, the 15- and 30-second TV commercials typical in the United States have shots 1–2 s in length, and Hollywood films feature shots that are on average 5 s in length [58,59]. In this study, the lengths of the snippets were uniformly distributed between two and 4 s, so each SV consisted of nine to eleven snippets with no temporal gaps in between.

The original video clips were randomly taken from two datasets: CRCNS-ORIG [60] and DIEM [61]. Both datasets include different styles of program that are commonly watched on a daily basis, such as documentaries, video games, sports events, movie trailers, advertisements, and TV news.

4.4. Data preprocessing

The raw eye-position data were segmented into blink, saccade, and fixation (or smooth-pursuit) periods. First, we extracted the blink periods by using the eyelid occlusion of both eyes. Specifically, we considered a blink to have occurred when both pupil diameters were zero for at least 100 ms (6 frames). Apart from the blink periods, artifacts detected by the eye tracker were removed by using a linear interpolation algorithm.

A standard method of identifying saccade and fixation periods is to detect saccades by using velocity and/or acceleration thresholds [46,25]. However, this method is only reliable at high sampling rates, such as 240 Hz. Instead, we used the mean-shift clustering method in the spatio-temporal domain, which is used in eye trackers with relatively low sampling rates, such as 30 and 60 Hz [40].

5. Results

5.1. Expt1: mental fatigue induced by cognitive tasks

We first determined whether or not the cognitive tasks mPASAT succeeded in inducing mental fatigue in the participants. To do so, we investigated subjective ratings and eye-tracking measures that have

Table 1
Fatigue-detection-model performance in terms of their average scores after 20 iterations of ten-fold cross-validation. F_{pre}: three feature sets related to oculomotor, blinks, and pupil measurements used in the previous studies, F_{sal}: saliency-based features, F_{emd}: features related to eye movement directions, and F_{ga}: features related to gaze allocation.

Model	Detection performance (%)			
	Accuracy	Precision	Recall	F-measure
F_{pre}	77.1	78.6	72.9	75.6
$F_{pre} + F_{sal}$	80.7	79.4	83.0	81.0
$F_{pre} + F_{emd}$	82.9	83.2	82.4	82.7
$F_{pre} + F_{ga}$	84.7	84.6	84.9	84.7
$F_{pre} + F_{sal} + F_{emd} + F_{ga}$	**91.0**	**91.4**	**90.3**	**90.8**

Values in bold in each column represent the highest of the performance.

been widely used as fatigue-correlated measures in previous studies. Next, we evaluated if our model could detect mental fatigue by cross-validation methods. We also tried to determine whether the three novel features sets could improve detection performance.

We investigated the participants' subjective ratings of mental fatigue before and after performing the mPASAT (Fig. 4A). Compared with the subjective ratings in phase 1, i.e., before engaging in the cognitive task, 12 and 14 out of 18 participants in phases 2 and 3, respectively, reported increased mental fatigue. A repeated-measures Friedman non-parametric ANOVA was done on the subjective ratings to compare phases 1, 2, and 3, and it found a statistically significant effect of phase ($\chi^2 = 7.36$, $n = 18$, d. f. $= 2$, $p < .05$). Dunn's multiple comparisons post-hoc test showed a significant increase in subjective ratings of mental fatigue from phases 1 to 3 ($p < .05$), but no significant difference between phases 1 and 2 or between phases 2 and 3.

Next, we investigated changes over the phases in eye-tracking measures that have been suggested as fatigue-correlated measures [24,25,18]. Specifically, we investigated pupil diameters, blinking, and saccade mean velocity. A one-way repeated measures ANOVA with post hoc Bonferroni multiple comparisons was used to calculate the statistical significance of the changes over the phases. In this analysis, we computed these measures of each participant by taking averages during each 5-minute phase.

We found a significant difference in both pupil diameters among phases ($F(2, 34) = 7.41$, $p < .005$ for the left eye; $F(2, 34) = 6.24$, $p < .005$ for the right eye, Fig. 4B). Post-hoc analyses revealed a significant decrease from phase 1 to 2 and 3 ($p < .05$, $p < .005$, respectively) for the left eye and from phase 1 to 3 ($p < .005$, Fig. 4B; from phase 1 to 2, $p = .14$) for the right eye. As for the blink behaviors, we found significant increases in duration, blink rate, and blink duration per minute over the phases. Among them, the blink duration per minute showed the biggest difference over the phases in terms of effect size ($F(2, 34) = 9.08$, $p < .001$, $\eta_p^2 = .348$, Fig. 4C). Post-hoc comparisons revealed significant increases along with the progression of phases (from phase 1 to 2, $p < .05$; from phase 1 to 3, $p < .001$). Finally, we analyzed changes in the measures of saccade mean velocity. Although they slightly decreased over the phases, there was no significant difference ($F(2, 34) = 3.14$, $p = .0562$, Fig. 4D).

Through this analysis, we found that the measures of pupil diameter and blinking as well as the subjective ratings showed significant changes from phase 1 to phase 3. Saccade mean velocity also showed a small decrease after performing cognitive tasks, although this change was not statistically significant. These changes in the eye-tracking measures are consistent with the results of previous studies that investigated the changes in these measures with increasing mental fatigue during a cognitive task [24,25,18]. Taken together, the results of the subjective and eye-tracking measures indicate that the participants experienced increased mental fatigue after engaging in the cognitive tasks two times, i.e., in phase 3. We thus regarded phase 1 as a non-fatigued circumstance and phase 3 as a fatigued circumstance, and proceeded to develop a model that classified the states of phases 1 and 3.

In addition, we investigated the changes in saliency-based features over the phases. Our hypothesis was that reflexive eye movement guided by bottom-up attention increases with mental fatigue, which could be captured as an increase in AUC scores of the saliency model. In our experiment, we found that the AUC scores significantly increased from phase 1 to 2 and 3 (A one-way repeated measures ANOVA with post hoc Bonferroni multiple comparisons, $F(2, 34) = 10.1$, $p < .001$). This result supports our hypothesis.

We built a fatigue detection model to differentiate eye-tracking data before and after performing the cognitive tasks. Specifically, we used eye-tracking data of 18 participants in phases 1 and 3. Our model used features extracted from each 30 s worth of eye tracking data in each SV. In other words, our model made a decision from 30 s of eye-tracking data as to whether a participant was in a fatigued or non-fatigued state. Phase 1 and 3 each consisted of nine 30-second SVs. Thus, the number of samples was $18 \times 9 \times 2 = 324$ (162 samples for non-fatigue states, and 162 samples for fatigued states; balanced datasets).

As a result of 20 iterations of ten-fold cross-validation, our model detected mental fatigue with 91.0% accuracy (91.4% precision, 90.3% recall, 90.8% F-measure, and chance 50%, Table 1). The feature selection process selected 55 of the 181 features as the most discriminative for classifiers for detecting mental fatigue. Although the selected features changed somewhat in accordance with the training dataset because our feature sets contained highly correlated features, they were selected from all six groups. We also evaluated our model by leave-one-subject-out cross-validation, where the models were trained by using the data collected from all of the participants except one and then tested on the data of the one participant left out of the training data set. We repeated this process for all participants, and obtained an accuracy of 88.6% (88.9% for younger adults, 88.2% for older adults who were more than 50 years of age).

Our model has uniqueness in its novel use of feature sets associated with gaze allocation, eye movement directions, and saliency-based metrics to better capture mental fatigue in natural viewing situations. Next, we investigated the contributions of these feature sets to fatigue detection performance. Specifically, we built models using a subset of the six feature sets and compared their model performances in ten-fold cross-validation. We did the feature selection and hyper-parameter optimization in the same way as in our model. First, we built a model using only the three feature sets based on previous studies, i.e., feature sets related to pupil measures, blinking, and oculomotor-based metrics. The eye-tracking measures used as fatigue-correlated measures in the previous subsection were included in the three feature sets. The model performance was 77.1% accuracy, which was 13.9% lower than that of our model. Next, we separately added each feature set to this model. Consequently, the model accuracies increased to 84.7% as a result of adding features related to gaze location, to 82.9% as a result of adding features related to eye movement directions, and to 80.7% as a result of adding saliency-based features. Therefore, we found that the three feature sets each improved the model's performance, and when taken together, they improved the model's performance by up to 13.9% (from 77.1 to 91.0%).

5.2. Expt2: sequence effect of repetitive video watching

We first investigated subjective ratings and eye-tracking measures that have been widely used as fatigue-correlated measures in previous studies. In the same manner as the first experiment, we performed a statistical analysis using a repeated-measures Friedman non-parametric ANOVA for subjective ratings and a one-way repeated measures ANOVA for eye-tracking measures including pupil diameters, blinking, and saccade mean velocity. As a result, we found no significant difference in the subjective ratings and eye-tracking measures over the phases (p > .05). Thus, we regarded the participants' state as being non-fatigued in phases 1 and 3.

We next extracted features from the eye-tracking data in phases 1 and 3 and investigated whether our model could estimate the participants' state as non-fatigued. Specifically, we obtained additional $11 \times 9 \times 2 = 198$ samples in this experiment. To make the description simple, we defined $D1_{nf}$ and $D3_{nf}$ to be the eye-tracking data of phases 1 and 3 in the experiment, respectively. We also defined $C1_{nf}$ and $C3_{f}$ to be eye-tracking data in phases 1 and 3 collected from the first experiment with the cognitive tasks. $D1_{nf}$, $D3_{nf}$, $C1_{nf}$ and $C3_{f}$ contained 99, 99, 162 and 162 samples, respectively. Here, $*_{f}$ and $*_{nf}$ represent data sets for fatigue and non-fatigue states, respectively.

First, we trained our mental fatigue detection model using $C1_{nf}$ and $C3_{f}$ and tested it using $D1_{nf}$ and $D3_{nf}$. As a result, our model estimated 175 out of 198 samples (88.3%) as non-fatigued states. Next, we used $C1_{nf}$, $C3_{f}$, and $D1_{nf}$ as training data and $D3_{nf}$ as test data. We found that 91 out of 99 samples (91.9%) were estimated as non-fatigued states. These results suggest that our fatigue-detection model captured changes resulting from mental fatigue induced by the cognitive tasks, not sequence effects of watching video.

6. Discussion

We built a model to detect mental fatigue of younger and older adults in natural viewing situations. We collected eye-tracking data from younger and older adults while they watched video and investigated whether our model could detect increased mental fatigue induced by the auditory cognitive tasks (mPASAT). In the first experiment, we used eye-tracking data before and after performing the cognitive tasks and showed that our model detected mental fatigue with 91.0% accuracy (chance 50%) from only 30 s worth of eye-tracking data. In our model, we added the three feature sets to better capture mental fatigue in natural viewing situations, although they have not been used before for detecting mental fatigue. We then evaluated the contributions of the three feature sets and found that they improved the model's performance by up to 13.9% (from 77.1% to 91.0%). In the second experiment where participants watched video clips without engaging in cognitive tasks, we found that our model could estimate the participants' state as non-fatigued with up to 91.9% accuracy. The results suggest that our model could capture changes in eye-tracking data resulting from mental fatigue induced by the cognitive tasks, not just sequence effects of repetitive video watching.

The results of the experiments indicated that mental fatigue affects the way people watch video clips in natural viewing situations and that our model could capture this change. In addition, the prediction power of the saliency model for fixations, or AUC scores, significantly increased after the participants engaged in the cognitive tasks, suggesting that eye movement might be more affected by bottom-up attention in a stimulus-driven manner when an individual is fatigued. Previous lab studies showed how neurodevelopmental and neurodegenerative diseases influence eye movements in similar natural viewing situations, and how a computational model could differentiate patients from healthy controls by capturing this difference [42,38,10,43]. In these studies, saliency-based features have been typically used as a means of characterizing the changes in the control of attention in natural viewing situations [42,10,43]. They explained these results on the basis of findings of neuroscience studies as follows. The control of attention and eye movements involves not only the visual area but also various other brain regions [62]. Neurodevelopmental and neurodegenerative diseases might affect their functions, which would lead to measurable changes in eye movement behaviors [62]. As for mental fatigue, previous psychological studies showed that it also affects multiple cognitive functions such as attention, motivation, and working memory [18]. However, whether and how mental fatigue affects eye movements in natural viewing situations remained largely uninvestigated. One of our contributions lies in providing the first empirical evidence of measurable eye-tracking signatures of mental fatigue in natural viewing situations, including saliency-based features.

The previous fatigue detection studies mainly investigated people younger than 50 years of age [29], despite that age-related changes in eye-tracking measures have been reported especially in older adults [29,32,41]. In our study, we experimentally found that measures used as fatigue-correlated measures such as saccade velocity, blinking, and pupil diameters changed between younger and older adults in a manner consistent with previous studies [29,32,41]. Some of these features such as blinking duration were not selected as a subset of model features by the feature selection process, and exhibited smaller changes associated with mental fatigue compared with their changes between age groups. In contrast, we found that some of the selected features such as saliency-based features exhibited relatively larger changes over the phases than the changes in age groups, which indicate that these features might be sensitive to mental fatigue and robust to target's age. The results showed that the combination of novel feature sets and feature selection method might be used to develop a fatigue-detection model robust to age-related changes in eye-tracking data.

In our experiment, we used auditory cognitive tasks (mPASAT) and confirmed whether or not the tasks induced mental fatigue in the participants by using subjective ratings as well as eye-tracking measures that have been widely used as fatigue-correlated measures in previous studies. About the subject ratings, 14 out of 18 participants reported increased mental fatigue after engaging in the cognitive tasks two times, which was a significant increase at the . 05 level. Interestingly, the other participants who did not report increased mental fatigue were more than 65 years old. Their eye-tracking measures still changed to imply an increase in mental fatigue: decreased pupil diameters and saccadic velocity, and longer and more frequent blinking. We cannot make any conclusions about the difference in the subjective ratings for mental fatigue between younger and older adults because the sample size of the participants was too small to perform a statistical comparison between age groups. If subjective and objective measures for mental fatigue tend to be different in older adults, a fatigue detection model might be important especially in the context of monitoring fatigue in older adults. In future work, we will collect data with more participants and compare the age-related differences in effect of mental fatigue on subjective ratings and eye-tracking measures.

Our work has several limitations. First, the small number of participants. We need to perform further research with more participants to strongly confirm our results. Second, we collected eye-tracking data in a lab setting. The controlled setting might influence the way people watch video clips. Third, the limited content of the video clips used in our experiment might be a potential problem. Although we made video clips by combining multiple styles of programming, we used only 10 min worth of eye-tracking data from each participant. There is a possibility that the high performance of our model might have resulted from over-fitting due to the limited content. Fourth, we treated a binary state of mental fatigue. A model capable of inferring mental fatigue on a scalar or ordinal scale would be more useful for monitoring and managing individual's health. This limitation mainly comes from the fact that we found significant changes in the subjective and eye-tracking measures only from phases 1 to 3. In future work, we need to collect data in which there are gradual increases in mental fatigue as the workload of the cognitive tasks increases; this could allow us to test whether or not our model could be extended to score mental fatigue on a scalar or ordinal scale.

7. Conclusion

In contrast to previous studies focusing on detecting mental fatigue during cognitive tasks, we aimed to develop a model enabling us to detect mental fatigue in natural-viewing situations when an individual is not performing cognitive tasks. In addition, considering the increasing demand for health monitoring for older adults, we also aimed to make our model robust to multiple age-related changes in eye-tracking measures. To this end, we devised a fatigue-detection model including (i) novel feature sets to better capture mental fatigue in natural-viewing situations and (ii) an automated feature selection method to select a feature subset enabling the model to be robust to the target's age. To test our model, we conducted the experiments and collected eye-tracking data from 29 younger and older participants while they watched video clips. Through the analyses, we demonstrated that our model could detect increased mental fatigue induced by the cognitive tasks from 30 s worth of eye-tracking data with 91.0% accuracy (chance 50%), which was up to 13.9% higher than a model based on the previous studies. We also showed that the prediction power of the saliency model for fixations significantly increased after the cognitive tasks, suggesting that eye movement in natural viewing situations might be more affected by bottom-up attention when an individual is fatigued. Moreover, we discussed the possibility that subjective and objective measures for mental fatigue tend to be different in older adults, and suggest that robustness of our model to age-related changes in eye-tracking measures would play a more important role in monitoring mental fatigue in a wider age group including older adults. Although we need to conduct further research, including an in-situ study, we believe that our results could help develop a model to monitor mental fatigue in everyday life.

Acknowledgments

This research was supported by the Japan Science and Technology Agency (JST) under the Strategic Promotion of Innovative Research and Development Program (this acknowledgement shall be retained in any derivative works that you publish).

References

[1] Alemdar H, Ersoy C. Wireless sensor networks for healthcare: a survey. Comput Netw 2010;54(15):2688–710.
[2] Favela J, Castro LA. Technology and aging. Aging research: methodological issues. Springer; 2015. p. 121–35.
[3] Yamada Y, Shinkawa K, Takase T, Kosugi A, Fukuda K, Kobayashi M. Monitoring daily physical conditions of older adults using acoustic features: a preliminary result. Stud Health Technol Inform 2018;247:301–5.
[4] Lu H, Frauendorfer D, Rabbi M, Mast MS, Chittaranjan GT, Campbell AT, et al. Stresssense: detecting stress in unconstrained acoustic environments using smartphones. Proc ACM Int Conf Ubiquitous Comput. 2012. p. 351–60.
[5] Lyu Y, Luo X, Zhou J, Yu C, Miao C, Wang T, et al. Measuring photoplethysmogram-based stress-induced vascular response index to assess cognitive load and stress. Proc SIGCHI Conf Hum Factor Comput Syst 2015:857–66.
[6] Pfleging B, Fekety DK, Schmidt A, Kun AL. A model relating pupil diameter to mental workload and lighting conditions. Proc SIGCHI Conf Hum Factor Comput Syst 2016:5776–88.
[7] Rahman T, Adams AT, Ravichandran RV, Zhang M, Patel SN, Kientz JA, et al. Dopple Sleep: a contactless unobtrusive sleep sensing system using short-range doppler radar. Proc ACM Int Conf Ubiquitous Comput 2015:39–50.
[8] Faurholt-Jepsen M, Busk J, Frost M, Vinberg M, Christensen E, Winther O, Bardram JE, Kessing L. Voice analysis as an objective state marker in bipolar disorder. Transl Psychiatry 2016;6(7):e856.
[9] Szatloczki G, Hoffmann I, Vincze V, Kalman J, Pakaski M. Speaking in Alzheimer's disease, is that an early sign? Importance of changes in language abilities in Alzheimer's disease. Front Aging Neurosci 2015;7.
[10] Zhang Y, Wilcockson T, Kim KI, Crawford T, Gellersen H, Sawyer P. Monitoring dementia with automatic eye movements analysis. Intell Dec Tech. 2016. p. 299–309.
[11] Shinkawa K, Yamada Y. Word repetition in separate conversations for detecting dementia: a preliminary evaluation on regular monitoring service. AMIA Summits on Translational Science Proceedings 2017 2018:206–15.
[12] Shinkawa K, Yamada Y. Topic repetition in conversations on different days as a sign of dementia. Stud Health Technol Inform 2018;247:641–5.
[13] Esposito A, Esposito AM, Likforman-Sulem L, Maldonato MN, Vinciarelli A. On the significance of speech pauses in depressive disorders: results on read and spontaneous narratives. Recent advances in nonlinear speech processing. Springer; 2016. p. 73–82.
[14] Tsanas A, Little MA, McSharry PE, Ramig LO. Accurate telemonitoring of Parkinson's disease progression by noninvasive speech tests. IEEE Trans Biomed Eng 2010;57(4):884–93.
[15] Boksem MA, Tops M. Mental fatigue: costs and benefits. Brain Res Rev 2008;59(1):125–39.
[16] Avlund K. Fatigue in older adults: an early indicator of the aging process? Aging Clin Exp Res 2010;22(2):100–15.
[17] Lorist MM, et al. Mental fatigue and task control: planning and preparation. Psychophysiology 2000;37(5):614–25.
[18] Hopstaken JF, Linden D, Bakker AB, Kompier MA. A multifaceted investigation of the link between mental fatigue and task disengagement. Psychophysiology 2015;52(3):305–15.
[19] Di Stasi LL, McCamy MB, Pannasch S, Renner R, Catena A, Canas JJ, et al. Effects of driving time on microsaccadic dynamics. Exp Brain Res 2015;233(2):599–605.
[20] Shahly V, Berglund PA, Coulouvrat C, Fitzgerald T, Hajak G, Roth T, et al. The associations of insomnia with costly workplace accidents and errors: results from the America insomnia survey. Arch Gen Psychiatry 2012;69(10):1054–63.
[21] Maghout-Juratli S, Janisse J, Schwartz K, Arnetz BB. The causal role of fatigue in the stress-perceived health relationship: a MetroNet study. J Am Board Fam Med 2010;23(2):212–9.
[22] Hardy SE, Studenski SA. Fatigue and function over 3 years among older adults. J Gerontol A Biol Sci Med Sci 2008;63(12):1389–92.
[23] Lin F, Chen D-G, Vance DE, Ball KK, Mapstone M. Longitudinal relationships between subjective fatigue, cognitive function, and everyday functioning in old age. Int Psychogeriatr 2013;25(02):275–85.
[24] Schleicher R, Galley N, Briest S, Galley L. Blinks and saccades as indicators of fatigue in sleepiness warnings: looking tired? Ergonomics 2008;51(7):982–1010.
[25] Di Stasi LL, Renner R, Catena A, Canas JJ, Velichkovsky BM, Pannasch S. Towards a driver fatigue test based on the saccadic main sequence: a partial validation by subjective report data. Transp Res Part C Emerg Technol 2012;21(1):122–33.
[26] Morad Y, Lemberg H, Yofe N, Dagan Y. Pupillography as an objective indicator of fatigue. Curr Eye Res 2000;21(1):535–42.
[27] Benedetto S, Pedrotti M, Minin L, Baccino T, Re A, Montanari R. Driver workload and eye blink duration. Transp Res Part F Traffic Psychol Behav 2011;14(3):199–208.
[28] Di Stasi LL, Catena A, Canas JJ, Macknik SL, Martinez-Conde S. Saccadic velocity as an arousal index in naturalistic tasks. Neurosci Biobehav Rev 2013;37(5):968–75.
[29] Dawson D, Searle AK, Paterson JL. Look before you (s)leep: evaluating the use of fatigue detection technologies within a fatigue risk management system for the road transport industry. Sleep Med Rev 2014;18(2):141–52.
[30] Azim T, Jaffar MA, Mirza AM. Fully automated real time fatigue detection of drivers through fuzzy expert systems. Appl Soft Comput 2014;18:25–38.
[31] Irving EL, Steinbach MJ, Lillakas L, Babu RJ, Hutchings N. Horizontal saccade dynamics across the human life span. Invest Ophthalmol Vis Sci 2006;47(6):2478–84.
[32] Hill RL, Dickinson A, Arnott JL, Gregor P, McIver L. Older web users' eye movements: experience counts. Proc SIGCHI Conf Hum Factor Comput Syst 2011:1151–60.
[33] Yamada Y, Kobayashi M. Detecting mental fatigue from eye-tracking data gathered while watching video. Proceedings of conference on artificial intelligence in medicine in Europe. 2017. p. 295–304.
[34] Ahlstrom C, et al. Fit-for-duty test for estimation of drivers' sleepiness level: eye movements improve the sleep/wake predictor. Transp Res Part C Emerg Technol 2013;26:20–32.
[35] Di Stasi LL, McCamy MB, Macknik SL, Mankin JA, Hooft N, Catena A. Saccadic eye movement metrics reflect surgical residents' fatigue. Ann Surg 2014;259(4):824–9.
[36] Craig A, Tran Y, Wijesuriya N, Nguyen H. Regional brain wave activity changes associated with fatigue. Psychophysiology 2012;49(4):574–82.
[37] Crawford B. The dependence of pupil size upon external light stimulus under static and variable conditions. Proc R Soc Lond B Biol Sci 1936;121(823):376–95.
[38] Crabb DP, Smith ND, Zhu H. What's on TV? Detecting age-related neurodegenerative eye disease using eye movement scanpaths. Front Aging Neurosci 2014;6.
[39] Di Stasi LL, McCamy MB, Catena A, Macknik SL, Canas JJ, Martinez-Conde S. Microsaccade and drift dynamics reflect mental fatigue. Eur J Neurosci 2013;38(3):2389–98.
[40] Xu P, et al. TurkerGaze: crowdsourcing saliency with webcam based eye tracking. arXiv preprint arXiv:1504.06755.
[41] Dowiasch S, Marx S, Einhäuser W, Bremmer F. Effects of aging on eye movements in the real world. Front Hum Neurosci 2015;9:46.
[42] Tseng P-H, Cameron IG, Pari G, Reynolds JN, Munoz DP, Itti L. High-throughput classification of clinical populations from natural viewing eye movements. J Neurol 2013;260(1):275–84.
[43] Wang S, Jiang M, Duchesne XM, Laugeson EA, Kennedy DP, Adolphs R, et al. Atypical visual saliency in autism spectrum disorder quantified through model-based eye tracking. Neuron 2015;88(3):604–16.
[44] Treisman AM, Gelade G. A feature-integration theory of attention. Cogn Psychol 1980;12(1):97–136.
[45] Itti L, Koch C, Niebur E. A model of saliency-based visual attention for rapid scene analysis. IEEE Trans Pattern Anal Mach Intell 1998(11):1254–9.
[46] Itti L. Quantifying the contribution of low-level saliency to human eye movements in dynamic scenes. Vis Cognit 2005;12(6):1093–123.
[47] Borji A, Itti L. State-of-the-art in visual attention modeling. IEEE Trans Pattern Anal Mach Intell 2013;35(1):185–207.
[48] Carmi R, Itti L. The role of memory in guiding attention during natural vision. J Vis 2006;6(9):4.
[49] Boser BE, Guyon IM, Vapnik VN. A training algorithm for optimal margin classifiers. Proceedings of the fifth annual workshop on computational learning theory 1992:144–52.
[50] Chang C-C, Lin C-J. LIBSVM: a library for support vector machines. ACM Trans Intell Syst Technol 2011;2(3):27.
[51] Guyon I, Weston J, Barnhill S, Vapnik V. Gene selection for cancer classification

using support vector machines. Mach Learn 2002;46(1–3):389–422.
[52] Yan K, Zhang D. Feature selection and analysis on correlated gas sensor data with recursive feature elimination. Sens Actuators B: Chem 2015;212:353–63.
[53] Huang MX, Kwok TC, Ngai G, Chan SC, Leong HV. Building a personalized, auto-calibrating eye tracker from user interactions. Proc SIGCHI Conf Hum Factor Comput Syst 2016:5169–79.
[54] Tass P, Rosenblum M, Weule J, Kurths J, Pikovsky A, Volkmann J, et al. Detection of n:m phase locking from noisy data: application to magnetoencephalography. Phys Rev Lett 1998;81(15):3291.
[55] Harel J, Koch C, Perona P. Graph-based visual saliency. Adv Neural Inf Process Syst 2006:545–52.
[56] Hjermstad MJ, Fayers PM, Haugen DF, Caraceni A, Hanks GW, Loge JH, et al. Studies comparing numerical rating scales, verbal rating scales, and visual analogue scales for assessment of pain intensity in adults: a systematic literature review. J Pain Symptom Manage 2011;41(6):1073–93.
[57] Cook DB, O'Connor PJ, Lange G, Steffener J. Functional neuroimaging correlates of mental fatigue induced by cognition among chronic fatigue syndrome patients and controls. Neuroimage 2007;36(1):108–22.
[58] Cutting JE, DeLong JE, Brunick KL. Visual activity in Hollywood film: 1935 to 2005 and beyond. Psychol Aesthet Creat Arts 2011;5(2):115–25.
[59] Bordwell D. Intensified continuity visual style in contemporary American film. Film Q 2002;55(3):16–28.
[60] Itti L, Carmi R. Eye-tracking data from human volunteers watching complex video stimuli. 2009http://crcns.org/data-sets/eye/eye-1/about.
[61] Mital PK, Smith TJ, Hill RL, Henderson JM. Clustering of gaze during dynamic scene viewing is predicted by motion. Cognit Comput 2011;3(1):5–24.
[62] Itti L. New eye-tracking techniques may revolutionize mental health screening. Neuron 2015;88(3):442–4.