MIND-WANDERING DETECTION MODEL WITH ELECTROENCEPHALOGRAM

Chutimon Rungsilp[1], Krerk Piromsopa[1,2], Atthaphon Viriyopase[3] and Kongpop U-Yen[3]
Chulalongkorn University, Bangkok, Thailand
[1]Department of Computer Engineering, Faculty of Engineering
[2]Research Group on Applied Digital Technology in Medicine
[3]Future Innovation Science and Technology

ABSTRACT

The study of mind-wandering is popular since it is linked to the emotional problems and working/learning performance. In terms of education, it impacts comprehension during learning which affects academic success. Therefore, we sought to develop a machine learning model for an embedded portable device that can categorize mind-wandering state to assist people in keeping track of their minds. We utilize a low-channel EEG to record the brain state and to build the predictive model because of its practicality and user-friendly. Most machine learning experiments in mind-wandering using EEG exhibit good individual-level performance. For the group-level technique, only a few research has developed a model. As a result, the goal of this research is to achieve a high-accuracy group-level model. Thus, Leave One Participant Out Cross Validation (LOPOCV) was used to assess the model correctness. This study shows that using a baseline normalization technique assists feature extraction and improves performance. The model was built using a support vector machine (SVM), and the best model achieved an accuracy value of 75.6 percent.

KEYWORDS

Mind-Wandering, Electroencephalogram, Support Vector Machine, Time-Frequency Analysis

1. INTRODUCTION

Mind-wandering is your thought that does not relate to current activity, commonly occurs in everyday life (Killingsworth and Gilbert, 2010). Your mind flew away to the past, future, or impossible occurrence while doing some tasks repeatedly. If you are mind-wandering during classroom instruction, you will miss out on lessons. This event is a negative impact on your retention and recognition. Some studies supported that it can cause faulty scenes and reduce working performance (Giambra, 1995) . Moreover, some research advocated that mind-wandering correlates with an emotional disorder, such as neuroticism, alexithymia, dissociative, and depression (Watts and Sharrock, 1985, Baer et al., 2006, Jensen et al., 2016). Thus, this study aimed to establish a machine learning model for an embedded portable device that can classify mind-wandering in real-time to help people monitor their minds.

To access mind-wandering sessions, two conceptual protocols are used to collect data. Firstly, it is called probe-caught sampling. This technique collects data by randomly sampling question subjects while doing a determined task repeatedly. The question is whether they are mind-wandering during an experiment in the previous ten minutes (Smallwood et al., 2007). This method has a drawback since people are intermittently aware of mind-wandering (Schooler et al., 2011). So, some sessions of mind-wandering were mistakenly reported (Pelagatti et al., 2020). To get rid of this issue, self-caught sampling was introduced. The self-caught sampling is a method that a subject performs a determined task repeatedly. If the subject is aware of mind-wandering during an experiment, report the mental state by pushing a button. This research used the self-caught sampling method to obtain dataset. The protocol is described in the next section.

After getting the protocol, we selected a biological instrument to record mind-wandering data. Hasenkamp (Hasenkamp et al., 2012) studied mind-wandering using functional magnetic resonance imaging (fMRI), which measures brain activity by monitoring variations associated with blood flow. There is a fluctuation of brain function between focus and mind-wandering during focus-attention meditation with breath counting.

Furthermore, pupil diameter and electroencephalogram (EEG) can detect mind-wandering (Pelagatti et al., 2020, Dhindsa et al., 2019). This study has chosen a few-channel EEG device, which measures electrical activity on the scalp. The reasons are portable usability function and affordable price.

There are several studies on mind-wandering using EEG. Smallwood and Beach et al. studied the effect of mind-wandering on time series EEG signals. They found the different amplitude of the P300 ERP component, which represents brain response of sensory, cognitive, or motor event (Kam et al., 2012). Braboszcz and Delorme discovered that the mind-wandering effect on EEG power frequency bands. Theta (4-7 Hz) and delta band (2-3.5 Hz) are increasing, but alpha (9-11 Hz) and beta band (15-30 Hz) are decreasing (Braboszcz and Delorme, 2011). This study applied machine learning to the EEG dataset. Due to the limitation of the device's resources, we selected the support vector machine (SVM) model. Moreover, SVM is suitable for real-time systems (Anh et al., 2012) and is the best classification performance (Lotte et al., 2007).

2. MATERIAL AND METHOD

This section describes research procedures in five parts, such as data acquisition, data cleansing, feature selection, model classification, and performance evaluation. The explanation is below.

2.1 Subject and Procedure

2.1.1 Subjects

We acquired the data from Future Innovation Research Science and Technology (FIRST) (Viriyopase, 2021, in preparation), which contains datasets from 22 right-handed subjects (male =12, female =10). They got Chulalongkorn University poster information. The subjects are 42.64(SD=7.39) years old on average, without any history of psychological disorder. This is confirmed by a PHQ-9 score below or equal to 4 (Kroenke et al., 2001). The subjects must give informed consent.

2.1.2 Procedure

Figure 1. Muse EEG recording device and its electrode positions (Liu et al., 2020)

Recording EEG Data, we used InteraXon MUSE 2016. The signals originate from AF7 and AF8 silver active electrode, and TP9 and TP10 conductive-silicon-rubber active electrode (referenced position with UI 10-10 system)(Jurcak et al., 2007). The instrument and the electrode position are shown in Figure 1. The EEG signal has a 256-Hz sampling rate and 12-bits scale resolution. The protocol for getting data is shown in Figure 2 and can be described as follows.

The data collecting protocol is modified from Braboszcz's study (Braboszcz and Delorme, 2011). A participant sits down, relaxes, and closes his/her eyes for 2 minutes before starting an experiment. The participant pushes a button to start the trial, then the start of trial session sound beeps randomly between 5-10 seconds. After the participants hear the sound, focus on the breath by counting 1 to 10 repeatedly. If he/she is aware that he/she does not concentrate on counting-breath during the trial, presses the button to report a mind-wandering and finished the trial session. All subjects repeatedly experiment for no more than 3 hours to prevent fatigue.
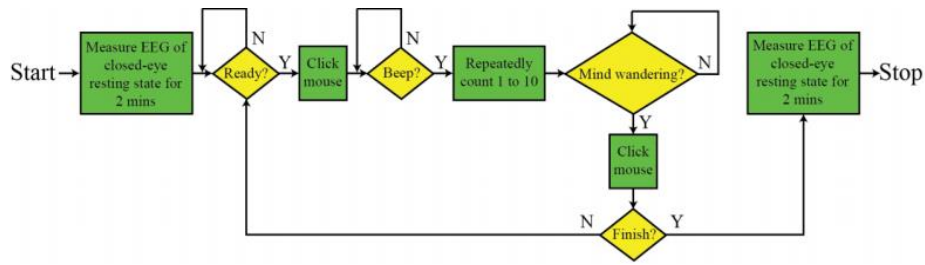
Figure 2. The EEG data acquisition protocol (Viriyopase, 2021, in preparation)

2.1.3 Data Structure

The data of each experiment contains 4-channel EEG time-series data (TP9, AF7, AF8, TP10) with a 512 Hz sampling rate. The data is recorded from EEG between starting session and reporting end session, which is shown in figure 3.
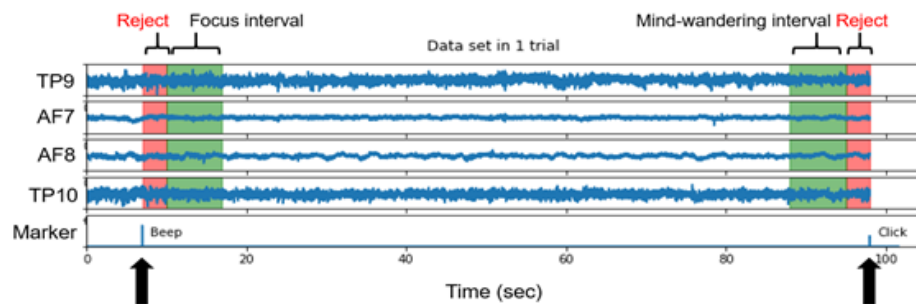

Figure 3. The dataset of each experiment

2.2 Data Cleansing

We considered that the subject was focus (FS) on the breath for 2 to 10 minutes after the short beep occurring. Mind-wandering (MW) is a duration between -10 to -2 minutes before pushing the report. We applied a bandwidth filtering of 0.5-30 Hz and rejected data with the artifact by the following criteria. The average value of electrode TP9 and TP10 with 50ms time window was gradients with greater than 10 µV/ms or difference of more than 100 µV (Krigolson et al., 2017). Finally, we got 990 cleaned interval data from 22 subjects (495 FS and 495 MW samples) which are usable for creating the model.

2.3 Feature Selection

We implemented time-frequency analysis and baseline normalization to create a feature vector of each data sample. Firstly, we select Morlet wavelet convolution in time frequency because of low computational complexity (Cohen, 2019). A complex Morlet wavelte equation is

$$w = e^{2i\pi ft}e^{\frac{-t^2}{2\sigma^2}},$$

when w is a product between a complex sine wave and a Gaussian window. i is an imaginary number. f is the frequency in Hz. t is time in seconds (define t from -2 to +2 seconds). σ is the width of the Gaussian, which are defined as:

$$\sigma = \frac{n}{2\pi f},$$

when n is a number of cycles, which affect the precision of temporal resolution and frequency resolution (define n is 1.5 to 8.0 round at frequency 3 to 30 Hz with linearly increasing (Braboszcz and Delorme, 2011)).

After we acquired time-frequency sample data, it has a problem that the power in all frequency does not have the same scale because of 1/f phenomenal and the difference of person, electrode, and trial. Baseline

normalization is used to get rid of this issue. Secondly, we applied three baseline normalization methods (Cohen, 2014).

• Decibel conversion (dB_{tf}) is the ratio between the power activity of each frequency and the baseline power activity of each frequency. The equation is

$$dB_{tf} = 10log10\left(\frac{activity_{tf}}{\overline{baseline_f}}\right),$$

when $\overline{baseline_f}$ is an average of the power activity of each frequency in baseline time. $activity_{tf}$ is power activity of frequency f and time t.

• Percentage change and baseline division ($prctchange_{tf}$) is a percent of the power activity relate to baseline. The equation is

$$prctchange_{tf} = 100\frac{activity_{tf}-\overline{baseline_f}}{\overline{baseline_f}}.$$

• Z-transform (Z_{tf}) is power activity scaling in the standard deviation unit of the baseline period. The equation is

$$Z_{tf} = \frac{activity_{tf}-\overline{baseline_f}}{\sqrt{n^{-1}\sum_{i=1}^{n}\left(baseline_{if}-\overline{baseline_f}\right)^2}},$$

when $baseline_{if}$ is the power activity of baseline at frequency f and time i. This study employs two minutes of closed eyelids before beginning an experiment as a baseline period.

The normalized sample was calculated from the mean power values in five frequency bands with time epoch of 2, 4, 6, and 8 seconds.

Eventually, we got five-frequency-band features (size of matrix is 5 features x 990 samples) with varying four-electrode positions, four-time-window size, and three baseline normalization techniques conditions. We would find the best condition for modeling in the performance evaluation topic.

2.4 Model Classification

We select the support vector machine (SVM). This model is a supervised learning model which construct an optimal separating hyperplane for classification. The dataset (D) for the training model contains five features and 990 samples with the defining condition of electrode positions, time window size, and baseline normalization techniques.

$$D = \{(\vec{x}_i, y_i); i = 1,2,...,990\},$$

when $\vec{x}_i = [x_{i1}, x_{i2}, ..., x_{i5}] \in R^5$ is a feature vector of sample i[th]. $y_i \epsilon \{1, -1\}$ is a class label.

The SVM decision function (f) can be defined as follow.

$$f(x) = sign(\vec{w}.K(\vec{x},\vec{z}) + b),$$

when b is a bias. $\vec{w}$ is the weight vector of the feature vector after converting with a kernel function ($K(\vec{x},\vec{z})$), low-dimensional transforms into the high-dimensional features. We select the linear kernel and radial basis function kernel (RBF), which give high performance in previous BCI application studies (Lotte et al., 2007).

Linear kernel can be defined as

$$K(\vec{x},\vec{z}) = \vec{x}.I,$$

when I is an identity matrix.

RBF kernel can be defined as

$$K(\vec{x},\vec{z}) = [e^{-\gamma\|\vec{x}_1-\vec{z}_1\|^2}, e^{-\gamma\|\vec{x}_2-\vec{z}_2\|^2}, ..., e^{-\gamma\|\vec{x}_u-\vec{z}_u\|^2}],$$

when $\vec{z} = [\vec{z}_1, \vec{z}_2, \vec{z}_3, ..., \vec{z}_u]$ is the vector of landmark, assist in feature dimension transformation (5-D to u-D). γ is a parameter that regulates model simplicity.

The optimal separating hyperplane is to maximize the margin of the parallel line. Each line is closed to the support vectors of different class. The margin maximizes when we minimize the following:

$$Minimize \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{N}\xi_i \,,$$

when $(w^T K(\vec{x},\vec{z}) + b) + \xi_i - 1 \geq 0$, $\xi_i \geq 0, i = 1,2,\dots,N$. ξ_i is an error of the sample i^{th}. C is a parameter that is a trade-off between error smallness and model simplicity.

We used a grid search to find best C and γ parameters. This searching technique examines all parameter competitors to get the best performance of the model.

2.5 Performance Evaluation

The studies of machine learning in mind-wandering using EEG mostly achieved high performance in the individual-level analysis (Kawashima and Kumano, 2017, Jin et al., 2019, Dhindsa et al., 2019). Few studies establish a model in the group-level method (Jin et al., 2020). Thus, this study aims to achieve a group-level model with high accuracy score. To accomplish this goal, we suppose that the baseline normalization technique will directly influence the predictive models.

We practiced leave-one-participant-out-cross-validation (LOPOCV) to evaluate the performance of all models with various conditions and parameters. LOPOCV provides a test/train dataset, leave one subject out from the total dataset as testing data and the other 21 subjects as training data. The cross-validation iteratively operates until each participant is test data once. The averaged accuracy values of all cross-validation indicate the model performance. The model had conditions that varied in time window lengths, baseline normalization methods, electrode positions, parameters, and kernel types. The result will discuss in the next section.

3. RESULT

We implement LOPOCV to the model by varying condition on the 990 samples (495 FS and 495 MW samples). We design the study in 3 parts as follows:

3.1 Performance Comparison of Varied Electrode's Position

We used data with a time window length of 8 seconds. The data is not normalized. We create the search gride model of each electrode's position and compare them.
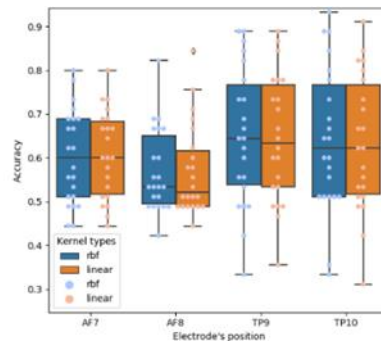

Figure 4. The model cross-validation accuracy values of each electrode and kernel type (without baseline normalization, time window length = 8 seconds)

This collection of boxplots in figure 4 depicts the accuracy distribution of both SVM kernels and the electrode placements of four electrodes. AF7, AF8, TP9, and TP10 have the average linear of SVM accuracy values of 0.61, 0.57, 0.63, and 0.64, respectively, and have average non-linear SVM accuracy values of 0.60, 0.56, 0.64, and 0.65, respectively. For all kernel types, the results show that the electrode positions of TP9 and TP10 perform better than AF7 and AF8.

3.2 Performance Comparison of Varied Baseline Normalization

To study the effect of baseline normalization with the predictive search gride models, we used data with a time window length of 8 seconds, electrode's position (TP9, TP10), and varied baseline normalization types.
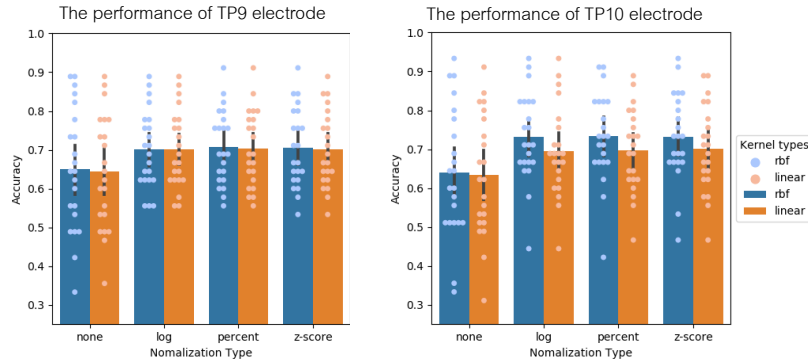

Figure 5. The model cross-validation accuracy values of each baseline normalization method and kernel type (time window length = 8 seconds). The left side is the performance of the TP9 electrode, and the right side is the performance of the TP10 electrode

The performance of each baseline normalization method of TP9 and TP10 location in Figure 5 is shown in bar graphs. The results demonstrate that models constructed from data with normalized baseline perform better than those built from data without normalization in both electrode position TP9 and TP10.

3.3 Performance Comparison of Varied Time-Window Length, Electrode's Position, and Baseline Normalization

The goal of this study was to find the optimum SVM model condition and parameter. To build each model, we change the electrode's location, baseline normalization, and time-window length. The top ten models with the best performance are shown in table 1.

Table 1. The Accuracy, precision, recall, area under curve, baseline normalization, and model parameter of each combination electrodes (top ten best performance)

electrode	**Accuracy**	precision	recall	area under curve	Baseline	C	γ	$K(\vec{x}, \vec{z})$	win
TP10	**Avg = 0.756,** Sd = 0.104	Avg =0.803, Sd = 0.132	Avg = 0.685, Sd = 0.216	Avg = 0.756, Sd = 0.102	z-score	1000	1	rbf	4
TP9,TP10	**Avg = 0.755,** Sd = 0.107	Avg = 0.792, Sd = 0.139	Avg = 0.686, Sd = 0.235	Avg = 0.756, Sd = 0.106	log	100	0.125	rbf	4
AF7,TP10	**Avg = 0.752,** Sd = 0.103	Avg = 0.804, Sd = 0.129	Avg = 0.684, Sd = 0.223	Avg = 0.752, Sd = 0.102	log	100	1	rbf	4
TP10	**Avg = 0.747,** Sd = 0.100	Avg = 0.804, Sd = 0.141	Avg = 0.667, Sd = 0.202	Avg = 0.748, Sd = 0.098	percent	1000	1	rbf	4
TP10	**Avg = 0.745,** Sd = 0.095	Avg = 0.808, Sd = 0.136	Avg = 0.661, Sd =0.198	Avg = 0.746, Sd = 0.094	log	1000	1	rbf	4
TP10	**Avg = 0.743,** Sd = 0.111	Avg =0.806, Sd = 0.148	Avg = 0.651, Sd = 0.241	Avg = 0.744, Sd = 0.111	log	1000	1	rbf	6
TP9,TP10	**Avg = 0.742,** Sd = 0.100	Avg =0.797, Sd = 0.131	Avg = 0.670, Sd = 0.225	Avg = 0.744, Sd = 0.100	percent	100	0.1	rbf	4
TP10	**Avg = 0.742,** Sd = 0.117	Avg =0.790, Sd = 0.173	Avg = 0.659, Sd = 0.240	Avg = 0.743, Sd = 0.116	percent	1000	1	rbf	6
TP10	**Avg = 0.742,** Sd = 0.108	Avg =0.803, Sd = 0.134	Avg = 0.656, Sd = 0.249	Avg = 0.742, Sd = 0.107	z-score	1000	1	rbf	6
All	**Avg = 0.740,** Sd = 0.096	Avg = 0.785, Sd = 0.126	Avg = 0.684, Sd = 0.217	Avg = 0.741, Sd = 0.095	log	100	0.125	rbf	4

The values which indicate the model performance are represented with bold style.
Avg, Average; Sd, Standard division; log, Decibel conversion; percent, Percentage change and baseline division; z-score, Z-transform.

We vary electrodes condition with one, two, and every positions. The result shows that the best model constructs from the dataset of TP10 electrode, 4 seconds time window length, log baseline normalization, and non-linear kernel. The performance measure of the best model is in the Figure 6.
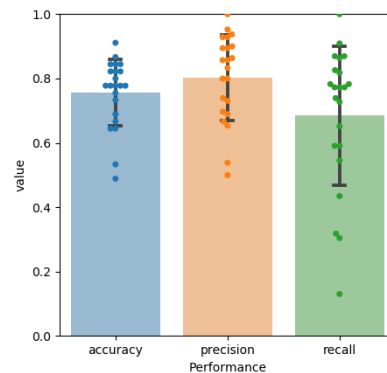

Figure 6. The cross-validation performance values of the best model conditions

In Figure 6, we presented the accuracy, precision, and recall of the best model. The model evaluates via LOPOCV. The dot plot overlay the bar is the performance of each validation round.

4. CONCLUSION

In this study, baseline normalization is employed for the first time in the feature extraction process for detecting Mw using low-channel EEG data. One advantage of this normalization technique is to enhance the feature for the group-level predictive model. Moreover, the predictive model constructs from low-channel EEG helps the device to convenient and easy to use. We applied LOPOCV to the 22-participant data and find the best conditions and parameters of the model. The study showed that the model with normalized baseline has higher performance than those without normalization. The best accuracy score is 75.6 % with 4-seconds time-window length, z-score baseline normalization, TP10 electrode's position, and non-linear SVM kernel.

Although our group-level analysis model can reveal the pattern of the brain associate with mind-wandering, the number of subjects is 22. In the future study, we would increase the number of participants to make the model more generalize. We hope that the application of this study helps people to observe their though. The students can take advantage of it to improve their cognition and retention. These are good for their learning or education.

REFERENCES

Anh, V. H., et al., 2012. A real-time model based support vector machine for emotion recognition through EEG. In: *2012 International Conference on Control, Automation and Information Sciences (ICCAIS).* IEEE, pp. 191-196.

Bare, R. A., et al. (2006) Using self-report assessment methods to explore facets of mindfulness. *Assessment,* Vol. 13, No. 1, pp 27-45.

Braboszcz, C. and Delorme, A., 2011. Lost in thoughts: neural markers of low alertness during mind wandering. *Neuroimage,* Vol. 54, No. 4, pp 3040-3047.

Cohen, M. X., 2014. *Analyzing neural time series data: theory and practice.* MIT press.

Cohen, M. X., 2019. A better way to define and describe Morlet wavelets for time-frequency analysis. *NeuroImage,* Vol. 199, pp 81-86.

Dhindsa, K., et al., 2019. Individualized pattern recognition for detecting mind wandering from EEG during live lectures. *PloS one,* Vol. 14, No. 9, pp e0222276.

Giambra, L. M., 1995. A laboratory method for investigating influences on switching attention to task-unrelated imagery and thought. *Consciousness and cognition,* Vol. 4, No. 1, pp 1-21.

Hasenkamp, W., et al., 2012. Mind wandering and attention during focused meditation: a fine-grained temporal analysis of fluctuating cognitive states. *Neuroimage,* Vol. 59, No. 1, pp 750-760.

Jensen, C. G., et al., 2016. General inattentiveness is a long-term reliable trait independently predictive of psychological health: Danish validation studies of the Mindful Attention Awareness Scale. *Psychological Assessment,* Vol. 28, No.5, pp e70.

Jin, C. Y., et al., 2019. Predicting task-general mind-wandering with EEG. *Cognitive, Affective, & Behavioral Neuroscience,* Vol. 19, No.4, pp 1059-1073.

Jin, C. Y., et al., 2020. Distinguishing vigilance decrement and low task demands from mind-wandering: A machine learning analysis of EEG. *European journal of neuroscience*.

Jurcak, V., et al., 2007. 10/20, 10/10, and 10/5 systems revisited: their validity as relative head-surface-based positioning systems. *Neuroimage,* Vol. 34, No. 4, pp 1600-1611.

Kam, J. W., et al., 2012. Mind wandering and motor control: off-task thinking disrupts the online adjustment of behavior. *Frontiers in human neuroscience,* Vol. 6, pp 329.

Kawashima, I. and Kumano, H., 2017. Prediction of mind-wandering with electroencephalogram and non-linear regression modeling. *Frontiers in human neuroscience,* Vol. 11, pp 365.

Killingsworth, M. A. and Gilbert, D. T., 2010. A wandering mind is an unhappy mind. *Science,* Vol. 330, No. 6006, pp 932-932.

Krigolson, O. E., et al., 2017. Choosing MUSE: Validation of a low-cost, portable EEG system for ERP research. *Frontiers in neuroscience,* Vol. 11, pp 109.

Kroenke, K., et al., 2001. The PHQ-9: validity of a brief depression severity measure. *Journal of general internal medicine,* Vol. 16, No. 9, pp 606-613.

Liu, R., et al., 2020. A Pilot Study on Electroencephalogram-based Evaluation of Visually Induced Motion Sickness. *Journal of Imaging Science and Technology,* Vol. 64, No.2, pp 20501-1-20501-10.

Lotte, F., et al., 2007. A review of classification algorithms for EEG-based brain–computer interfaces. *Journal of neural engineering,* Vol. 4, No. 2, pp R1.

Pelagatti, C., et al., 2020. A closer look at the timecourse of mind wandering: pupillary responses and behaviour. *PLoS one,* Vol. 15, No. 4, pp e0226792.

Schooler, J. W., et al., 2011. Meta-awareness, perceptual decoupling and the wandering mind. *Trends in cognitive sciences,* Vol. 15, No. 7, pp 319-326.

Smallwood, J., et al., 2007. The lights are on but no one's home: Meta-awareness and the decoupling of attention when the mind wanders. *Psychonomic bulletin & review,* Vol. 14, No. 3, pp 527-533.

Viriyopase, A., 2021(in preparation). Catch mind wandering by deep learning.

Watts, F. N. and Sharrock, R., 1985. Description and measurement of concentration problems in depressed patients. *Psychological Medicine,* Vol. 15, No. 2, pp 317-326.