

# Generalizability of EEG-based Mental Attention Modeling with Multiple Cognitive Tasks

Aung Aung Phyo Wai\*, Maokang Dou\* and Cuntai Guan\*\*

**Abstract**—Attention is the foundation of a person’s cognitive function. The attention level can be measured and quantified from the electroencephalogram (EEG). For the study of attention detection and quantification, we researchers usually ask the subjects to perform a cognitive test with distinct attentional and inattentional mental states. Different attention tasks are available in the literature, but there is no empirical evaluation to quantitatively compare the attention detection performance among the tasks. We designed an experiment with three typical cognitive tests: Stroop, Eriksen Flanker, and Psychomotor Vigilance Task (PVT), which are arranged in a random order in multiple trials. Data were collected from ten subjects. We used six standard band power features to classify the attention levels in four evaluation scenarios for both subject-specific and subject-independent cases. With cross-validation for the subject-independent case, we achieved a classification accuracy of 61.6%, 63.7% and 65.9% for PVT, Stroop and Flanker tasks respectively. We achieved the highest accuracy of 74.1% and 65.9% for the Flanker test in the subject-dependent and subject-independent cases respectively. Our evaluation shows no statistically significant differences in classification accuracy among the three distinct cognitive tasks. Our study highlights that EEG-based attention recognition can generalize across subjects and cognitive tasks.

**Keywords**— Brain-Computer Interface, Attention, Cognitive tests, Subject-independent cross-validation.

## I. INTRODUCTION

A Brain-Computer Interface (BCI) uses brain signals from a person enabling to communicate or control of external devices, and assess the person’s cognitive states [1]. Recently, BCI applications have been targeted to healthy people ranging from cognitive training, entertainment to medical evaluations etc. with miniaturization of electroencephalogram (EEG) sensing and user-friendly form-factor [2]. According to the differences in level of user’s control and external stimulus required or not, BCI can be classified into active, reactive and passive types [1]. The passive BCI derives its outputs from arbitrary brain activities without any voluntary control or mental tasks. That enables passive BCI simple to operate as mostly single or a few electrodes around frontal brain lobe is required to test user’s attention and mental states. With the improvements and development of dry wearable EEG headband, passive BCI can be incorporated in our daily life and multiple usage scenarios [3].

Passive BCI might capture cognitive states of a subject without distracting the subject from the primary task or affecting the subject because of the judgments of observers [3].

Aung Aung, Dou Maokang and Guan Cuntai are with the School of Computer Science and Engineering, Nanyang Technological University, 50 Nanyang Ave, Singapore 639798.

\*Authors have equal contributions. \*\*Cuntai Guan is the Corresponding author. Email: ctguan@ntu.edu.sg

In addition, with advantages of electroencephalogram (EEG), such as low cost, high availability, convenient setup, etc., EEG-based passive BCI has gained increasing popularity in mental states recognition researches [4]. Many studies have shown that brain activities in the frontal and pre-frontal lobe correspond to the subject’s personalities, emotions, attentions, etc. Therefore, EEG measured at frontal areas can assess and quantify the changes in one’s attention state [5].

Attention is important for and positively correlated with learning, memory and cognition capability. Attention affects the performance of everyone’s daily life that involved in most of the tasks we performed that can be measured and quantified using EEG. Thus, effective methods for measuring people’s attention level is useful in many applications, such as improving learning efficiency, vigilance detection while driving, attention training and rehabilitation [6]. However, attention is a complex neural and psychological phenomenon which involves various brain structures and mechanisms. We can broadly classify Attention into (1) selective attention with focusing on a specific object for a certain period despite distractions. (2) sustained attention for maintaining a focus on one object for a lengthy period. (3) divided attention to focus on multiple objects simultaneously and (4) executive attention for blocking out irrelevant features while focusing on important objects relevant to EEG-based mental attention measures. The large-scale randomized trials of attention deficit hyperactivity disorder (ADHD) children with BCI-based attention training showed both improvements ADHD symptoms and maintain the improvements after training [7]. Although different surrogate cognitive tests are used to model for classifying attention levels, there is lack of empirical study on performance difference between these cognitive tasks and generalizability of these tests in recognizing attention.

In this study, we design an experiment to test the attention recognition performance with three cognitive tests in subject-dependent and subject-independent cross validations. Section II explains the details of experiment design, evaluation and analysis methods. We present some discussion on analysis results and future works in Section III and conclusion in Section IV.

## II. MATERIALS AND METHODS

This study focuses on three cognitive tests that measures and evaluate selective attention and sustained attention types such as Stroop, Flanker, and psychomotor vigilance task (PVT) tests. Selective attention is important for intelligent

TABLE I  
CV OPTIONS USED IN PERFORMANCE EVALUATION OF DIFFERENT COGNITIVE TESTS

CV Option	Meaning	Description
L1BO	Leave-One-Block-Out	Similar to 9 folds. Single block contains individual (A, IA) task pair.
L1SessO	Leave-One-Session-Out	Similar to 3 folds. Single session includes all three blocks of (A, IA) pairs.
L1TO	Leave-One-Test-Out	Train on 2 tests and remaining test as Test data. We defined three variants according to the data segmentation.
L1SubO	Leave-One-Subject-Out	Train on N-1 subjects and remaining subject as Testing. Subject independent cross validation.

behaviours since people have to select task-relevant information and keep themselves from distracted by misleading information [8]. Sustained visual attention is critical for performing cognitive tasks such as driving, concentrating in lectures, and monitoring sophisticated equipment [9]. Stroop task assesses a subject’s selective attention where the subject is required to name the ink color of words [10]. Flanker task also tests a subject’s selective attention where the subject is presented with a target stimulus that is surrounded by non-target stimuli, and the subject is required to respond to the target stimulus only [11]. PVT measures the speed with which subjects respond to a visual stimulus [9]. To minimize rapid reduction of subjects’ attention level due to boredom and repetitiveness, we randomize the sequence of the above mentioned tasks in multiple sessions as shown in Figure 1. Each experiment session has three blocks of respective task as attention class (A) and the same condition for inattention class (IA) in all blocks. For IA class, we instructed the subject slowly looks around the screen border without focusing on specific position on the screen. In each attention (A) duration, subject need to repeat the same task about 10-13 trials according to subject’s response times. For whole inattention (IA) duration, subject will repeatedly perform the task until the task ends. Both A and IA tasks are approximately 30 s each in Figure 1(a). Each test subject follows on-screen instruction and stimulus presentation on both tasks and requires to press the space bar with their major arm to capture response time in different attention tasks.

blinks and incorrect gazes during the experiment. The pre-processing step includes the sample-based processing of input EEG signals,  $X \in R^{N \times M}$  where N is the number of channels and M is the number of EEG samples. We first apply zero-phase fourth order IIR band-pass filters of 0.5 to 45 Hz and normalize data. We then perform segment-based processing with The pre-processed EEG data firstly segmenting EEG data into 2 s epoch with overlapped of 100 ms window. Besides discarding trials with ocular artifacts using eye tracker’s gaze conditions, we further remove artifacts by applying a set of moving average filters. Finally, we extract six bandpower features ( $\delta(0.5 - 4Hz)$ ,  $\theta(4 - 8Hz)$ ,  $\alpha(8 - 12Hz)$ ,  $\beta_{low}(12 - 18Hz)$ ,  $\beta_{high}(18 - 30Hz)$ ,  $\gamma(30 - 40Hz)$ ) per channel by applying filter banks to compute the energy of the respective bands. With distinct attention (A) and inattention (IA) representative label as per protocol design 1, attention recognition is regarded as binary classification to classify attention and inattention resulting to provide attention scores that reflects user’s mental attention. We used SVM classifier with RBF kernel ( $C=2$ ,  $\gamma=0.5$ ) to train attention models with baseline attention and inattention class data as SVM often performed better than other classifiers [12], [13].

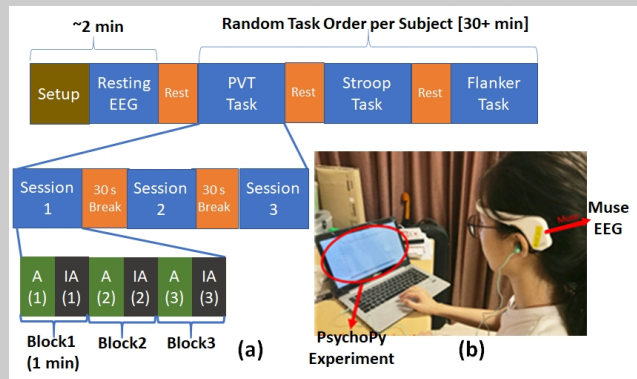


Fig. 1. Experiment Design and Protocol (a) Data collection scenario [A: attention, IA: inattention] (b) Subject performs the experiment with wearable EEG headband.

Figure 2 shows the processing pipeline to classify and recognize attention levels from EEG data. Eye tracker data are mainly used to remove the trials with excessive eye

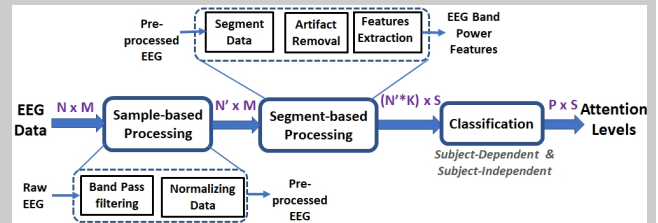


Fig. 2. Data analysis methodology and processing pipeline

### III. DATA ANALYSIS AND RESULTS

We got the written informed consent from all participants before the experiment. After setting up the experiment, 'resting EEG' data of closed eyes and open eyes are recorded before cognitive tests. We collected data from 10 healthy adults using wearable EEG Muse headband (Interaxon Inc.) and desktop eye tracker (Tobii AB) to capture subjects’ brain signals and eye gazes respectively. We use four electrodes positioned at Tp9, AF3/Fp1, AF4/Fp2 and Tp10 with common reference located at Fpz. For fair performance comparison, we used different cross validation options as shown in Table I to evaluate the performance of attention classification among different tasks while using the same analysis pipeline. The subject-dependent validation includes L1BO, L1SessO and

LITO whereas L1SubO is subject-independent validation. LITO can be divided into three variants according to number of data blocks included in evaluation. LITO-Block test data include single block, LITO-Sess test data include three blocks and LITO-Sub test data include nine blocks (all sessions data). In each LITO variant, we only use any two tests (out of PVT, Stroop and Flanker) as training data and the remaining as testing data.

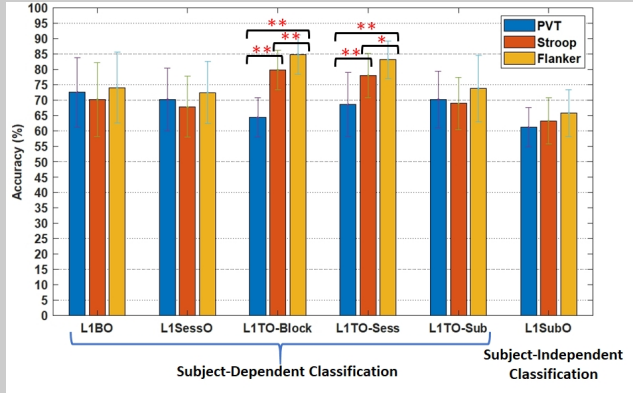


Fig. 3. Performance Comparison between Subject-Dependent and Subject-Independent classification accuracy.

Figure 3 shows accuracy comparison of three cognitive tests in different cross validation options of Table I. Although being a small sample size, accuracy difference between subject-dependent without LITO variants and subject-independent validations are less than 10%. This highlights attention modeling using three cognition tests can represent the general attention states of the subjects. There is no statistical significant difference among different task pairs (PVT, Stroop, Flanker) in accuracy except L1TO-Block and L1TO-Sess as shown in Figure 3 using Wilcoxon Rank Sum test.

Although no significant difference in accuracy among cognitive tests in subject-independent evaluation, there is a statistically significant difference in a comparison of behavioral response in terms of reaction time as shown in Figure 4. As expected, less significant difference between Stroop and Flanker tests as both represent surrogate measures of selective attention. Although PVT test is shorter reaction time than other two tests, its attention classification accuracy is lower than others although no statistically significant difference among them. This further validates generalizability of EEG-based attention modeling using those cognition tests to recognize spontaneous and continuous attention levels. Though no statistical significant differences, attention modeling with Flanker tests achieves 4–6% higher accuracy than other two tests as shown in Figure 4(b).

We use six EEG band power features in our evaluation analysis. Among these features, different combination of  $\theta$ ,  $\alpha$  and  $\beta$  band power features are used to derive focus or engagement index without modeling [14]. Also, the inclusion of  $\delta$  in features must be careful as different artifacts may contain in the EEG signals resulting biases towards potential artifacts

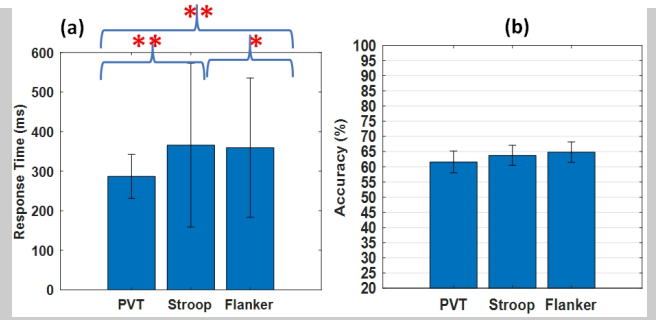


Fig. 4. Comparison between behavioral and EEG responses among cognitive tests (a) reaction time (b) subject-independent accuracy [\*\* indicates significant difference at  $p < 0.001$  and \* indicates significant difference at  $p < 0.05$ ].

or noises. With dry EEG electrodes, higher frequency like  $\gamma$  band is susceptible to external noises and interference. But we consider all six standard EEG band powers to further evaluate whether each band power performs differently between different cognition tests or not in attention modeling. So we evaluate whether significant difference with specific band power features between different test pairs in terms of percentage of subjects as shown in Table. II. Interestingly, only  $\beta_{Low}$  band power features show no statistical significant difference among tests. This highlights the similar  $\beta_{Low}$  features representation are captured among cognition tasks. The  $\theta$  band power features show high significant difference between test pairs that concludes high variability of this feature among subjects in selected cognitive tasks.

TABLE II

STATISTICAL SIGNIFICANT DIFFERENCE IN BAND POWER FEATURES COMPARISON BETWEEN DIFFERENCE PAIRS OF COGNITIVE TESTS. [% SHOW NUMBER OF SUBJECTS WITH STATISTICAL SIGNIFICANT DIFFERENCES AT  $p < 0.001$  LEVEL.

Tests/Bands	$\delta$	$\theta$	$\alpha$	$\beta_L$	$\beta_H$	$\gamma$
PVT Vs Stroop	10 %	30 %	0	0	10 %	0
PVT Vs Flanker	0	40 %	0	0	20 %	0
Flanker Vs Stroop	0	20 %	20 %	0	0	10 %

Figure 5 shows the accuracy comparison between subject-dependent (L1BO) and subject-independent (L1SubO) evaluations of Flanker test in all subjects. Except subject 'S9', the rest of the subject shows accuracy in L1BO is higher than that of L1SubO. This result highlights to further conduct in-depth analysis of how features from 'S9' differs with other subjects. The Wilcoxon Rank Sum test with the mean subject accuracy shows marginally statistical significant difference at  $p = 0.043$  between L1BO and L1SubO. The lower in accuracy with L1SubO of 65.9% can possibly be due to the small sample sizes of 10 subjects as well as no optimal features selection. Related to results in Table. II, subject-specific features selection and optimization might require for performance improvements due to differences and similarity in band power feature representations among subjects.

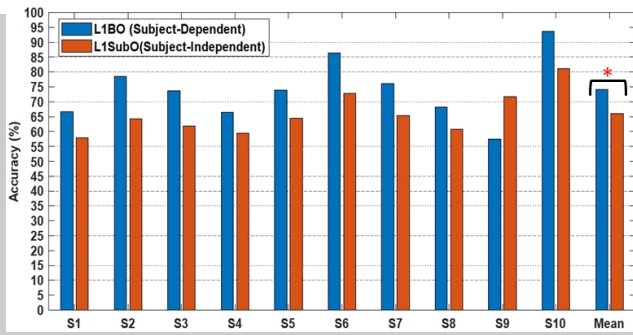


Fig. 5. Performance Comparison between subject-dependent and subject-independent classification using flanker test.

#### IV. DISCUSSION AND CONCLUSION

Our study shows the similarity in attention recognition performance among different cognitive tests. But the significant accuracy differences among tests in L1TO-Block and L1TO-Sess (Figure 3) might be due to combination of two tests instead of single test like L1SessO. According to features differences among subjects in each feature, we can highlight  $\theta$  band features are highly inconsistent features among subjects. Although we did quantitative comparison using 4-channel EEG, there is still a debate on the quality of measurements taken from consumer EEG headbands [15]. We need to further validate our findings to simultaneously take measurements from research-grade wet EEG and consumer-grade dry EEG with in-depth data analysis and systematic performance comparison [13], [16].

Still, our results only consider the accuracy criteria with standard EEG band power features. We still need to further test whether similar findings can achieve with other EEG features and more numbers of subjects. We will further test on collecting data from more subjects and applying no manually crafted features by applying deep learning approaches [17]. We will investigate on other temporal, spectral and spatial features beyond basic band power features to model attention. This study highlights the generalizability of EEG-based mental attention models using different cognitive tasks. With the prevalence of the consumer wearable headband, the applications of passive BCI to measure and quantify user's cognitive health and task improvements can be seen in real-world usages in near future [18], [3].

#### ACKNOWLEDGMENT

The study was reviewed and approved by the Institutional Review Board at Nanyang Technological University under reference number, IRB-2018-12-011. We would like to thank all subjects for their participation in our experiment.

#### REFERENCES

[1] L. F. Nicolas-Alonso and J. Gomez-Gil, "Brain computer interfaces, a review," *Sensors (Basel, Switzerland)*, vol. 12, no. 2, pp. 1211–1279, 2012, 22438708[pmid]. [Online]. Available: <https://www.ncbi.nlm.nih.gov/pubmed/22438708>

[2] A. A. P. Wai, H. Dajiang, and N. S. Huat, "Iot-enabled multimodal sensing headwear system," in *2018 IEEE 4th World Forum on Internet of Things (WF-IoT)*, 2018, pp. 286–290. [Online]. Available: <https://app.dimensions.ai/details/publication/pub.1103854459>

[3] P. Aricò, G. Borghini, G. D. Flumeri, N. Sciaraffa, and F. Babiloni, "Passive BCI beyond the lab: current trends and future directions," *Physiological Measurement*, vol. 39, no. 8, p. 08TR02, aug 2018.

[4] H.-S. Chiang, K.-L. Hsiao, and L.-C. Liu, "Eeg-based detection model for evaluating and improving learning attention," *Journal of Medical and Biological Engineering*, vol. 38, no. 6, pp. 847–856, 2018. [Online]. Available: <https://doi.org/10.1007/s40846-017-0344-z>

[5] N.-H. Liu, C.-Y. Chiang, and H.-C. Chu, "Recognizing the degree of human attention using eeg signals from mobile sensors," *Sensors (Basel, Switzerland)*, vol. 13, no. 8, pp. 10273–10286, Aug 2013. [Online]. Available: <https://www.ncbi.nlm.nih.gov/pubmed/23939584>

[6] L. Jiang, C. Guan, H. Zhang, C. Wang, and B. Jiang, "Brain computer interface based 3d game for attention training and rehabilitation," in *2011 6th IEEE Conference on Industrial Electronics and Applications*, 2011, pp. 124–127.

[7] C. G. Lim, X. W. W. Poh, S. S. D. Fung, C. Guan, D. Bautista, Y. B. Cheung, H. Zhang, S. N. Yeo, R. Krishnan, and T. S. Lee, "A randomized controlled trial of a brain-computer interface based attention training program for adhd," *PLoS ONE*, vol. 14, no. 5, pp. 1–16, 05 2019. [Online]. Available: <https://doi.org/10.1371/journal.pone.0216225>

[8] C. M. Jung, J. M. Ronda, C. A. Czeisler, and K. P. J. Wright, "Comparison of sustained attention assessed by auditory and visual psychomotor vigilance tasks prior to and during sleep deprivation," *Journal of sleep research*, vol. 20, no. 2, pp. 348–355, Jun 2011, 20819145[pmid]. [Online]. Available: <https://www.ncbi.nlm.nih.gov/pubmed/20819145>

[9] —, "Comparison of sustained attention assessed by auditory and visual psychomotor vigilance tasks prior to and during sleep deprivation," *Journal of sleep research*, vol. 20, no. 2, pp. 348–355, Jun 2011, 20819145[pmid]. [Online]. Available: <https://www.ncbi.nlm.nih.gov/pubmed/20819145>

[10] M. J. M. Lamers, A. Roelofs, and I. M. Rabeling-Keus, "Selective attention and response set in the stroop task," *Memory & Cognition*, vol. 38, no. 7, pp. 893–904, Oct 2010. [Online]. Available: <https://doi.org/10.3758/MC.38.7.893>

[11] T. J. McDerrott, A. I. Wiesman, A. L. Proskovec, E. Heinrichs-Graham, and T. W. Wilson, "Spatiotemporal oscillatory dynamics of visual selective attention during a flanker task," *NeuroImage*, vol. 156, pp. 277–285, Aug 2017, 28501539[pmid]. [Online]. Available: <https://www.ncbi.nlm.nih.gov/pubmed/28501539>

[12] C.-C. Chang and C.-J. Lin, "Libsvm: A library for support vector machines," *ACM Trans. Intell. Syst. Technol.*, vol. 2, no. 3, May 2011. [Online]. Available: <https://doi.org/10.1145/1961189.1961199>

[13] F. Lotte, L. Bougrain, A. Cichocki, M. Clerc, M. Congedo, A. Rakotomamonjy, and F. Yger, "A review of classification algorithms for EEG-based brain-computer interfaces: A 10-year update," *J. Neural Eng.*, vol. 15, no. 3, p. 031005, 2018. [Online]. Available: <http://stacks.iop.org/1741-2552/15/i=3/a=031005>

[14] V. Sandhu, A. A. P. Wai, and C. Y. Ho, "Evaluation of learning performance by quantifying user's engagement investigation through low-cost multi-modal sensors," in *2017 International Conference on Orange Technologies (ICOT)*, Dec 2017, pp. 180–183.

[15] E. Ratti, S. Waninger, C. Berka, G. Ruffini, and A. Verma, "Comparison of medical and consumer wireless eeg systems for use in clinical trials," *Frontiers in human neuroscience*, vol. 11, pp. 398–398, Aug 2017, 28824402[pmid]. [Online]. Available: <https://www.ncbi.nlm.nih.gov/pubmed/28824402>

[16] J. W. Y. Kam, S. Griffin, A. Shen, S. Patel, H. Hinrichs, H.-J. Heinze, L. Y. Deouell, and R. T. Knight, "Systematic comparison between a wireless eeg system with dry electrodes and a wired eeg system with wet electrodes," *NeuroImage*, vol. 184, pp. 119–129, Jan 2019, 30218769[pmid]. [Online]. Available: <https://www.ncbi.nlm.nih.gov/pubmed/30218769>

[17] A. Craik, Y. He, and J. L. Contreras-Vidal, "Deep learning for electroencephalogram (EEG) classification tasks: a review," *Journal of Neural Engineering*, vol. 16, no. 3, p. 031001, apr 2019.

[18] O. E. Krigolson, C. C. Williams, A. Norton, C. D. Hassall, and F. L. Colino, "Choosing muse: Validation of a low-cost, portable eeg system for erp research," *Frontiers in Neuroscience*, vol. 11, p. 109, 2017. [Online]. Available: <https://www.frontiersin.org/article/10.3389/fnins.2017.00109>