Discovery of Event-Related Potentials during a Cognitive Process of Comparison Operation

Keisuke Murai* , Hidetake Uwano* , Yoshiharu Ikutani † and Takatomi Kubo†
* National Institute of Technology, Nara College, Nara, Japan
E-mail: murai@info.nara-k.ac.jp, uwano@info.nara-k.ac.jp
† Nara Institute of Science and Technology, Nara, Japan
E-mail: ikutani.yoshiharu.ip8@is.naist.jp,takatomi-k@is.naist.jp

***Abstract*—Cognitive processes and mental states of software developers have been investigated using electroencephalogram (EEG) in software engineering. However, the feasibility of event-related potential (ERP) analyses for program comprehension were rarely examined. In this study, we aimed to discover the ERPs related to program comprehension. We conducted an experiment to capture ERPs with ten software developers during a cognitive process of comparison operation as one of the simple program comprehension processes. As results, ERP waveforms were observed in the EEG data of four subjects. In addition, we confirmed the inter-stimulus difference in peak amplitudes and peak latency.**

I. Introduction

Software developers make many decisions in their daily work ranging from naming variables to selecting architecture of overall system. As the decision-making processes in software development and maintenance are mostly dependent on developers' experience and intuition, it is difficult to identify the basis of decision [1], [2]. Therefore, despite its huge role in software development, we know little about how and when developers make decisions in their daily work. If we can identify when developers made a decision and what information he/she focused at the time, it might be possible to accumulate the evidences to elucidate their decision-making processes.

Recent studies have used biosignal measurements, e.g., eye movements and brain activities, to detect changes in cognitive processes and mental states of software developers [3], [4]. Electroencephalogram (EEG) reflects subjects' mental state with smaller costs and restrictions than those of other measurement methods of brain activities. In EEG data analyses, capturing event-related potentials (ERP) is one of the common methods. ERPs can be a potential indicator that reflect the timing when a cognitive process take place in the human brain [5].

However, the feasibility of ERP analyses for program comprehension are rarely examined. In this study, we aimed to discover the ERPs related to program comprehension. We conducted an experiment to capture ERPs with ten software developers during a cognitive process of comparison operation as one of the simple program comprehension processes. In addition, we examined the effects of inter-stimulus difference observed in ERP waveforms.

The primary contributions of this study are as followings:

1) We found the ERP waveforms in EEG data during a simple program comprehension task with comparison operation.
2) We found that ERP waveforms are different between conditional expression evaluated as True and False.

II. Related works

Recent studies that quantifying mental states of software developers using biosignals are attracting attention increasingly from software engineering researchers and practitioners. In particular, many studies have challenged to investigate cognitive processes of software developers during program comprehension, which consumes more than 50 percent of time in their daily work [6]. These studies can pave the way for future studies that systematically explore hypotheses to build theoretical foundations of program comprehension [7].

For instances, Lee et al. compared EEG data of expert programmers with those of novices during reading Java code snippets and demonstrated that experts' beta and gamma waves have higher amplitudes than those in novices [8]. Frits et al. combined three biosignals involving EEG, electrodemal activity, and eye movements to predict the difficulty of program comprehension tasks [3]. Müller et al. employed heart rate in addition to the previous three sensors and classified developers' mental states (i.e. stuck or in flow, happy or frustrated) with around 70 percent accuracy [4]. However, these prior EEG studies mainly focused on frequency components. ERPs in software engineering domain have rarely been argued.

ERPs are voltage fluctuations that are associated in time with some physical or mental occurrence [5]. In general, ERP analysis is a useful tool to investigate when subjects' mental processes occur because it examines a time course of EEG data on the milliseconds order. Kutas and Hillyard's early study demonstrated that N400 response (i.e. a negative voltage fluctuation observed around 400 msec after the event onset) appears when subjects face grammatical errors in a text [9]. Many following studies have associated the N400 response in

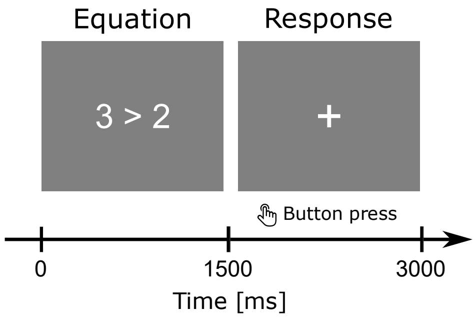


Fig. 1. Design of the conditional expression task

EEG data with various natural language processing and object recognition tasks (see [10] for a review). Further, Lau et al. combined the N400 responses in natural language processing with high spacial-resolution data obtained by functional magnetic resonance imaging (fMRI) and challenged to model the language network in the human brain [11]. This study implies that combinations of ERP and fMRI evidences, each of which covers high time- and spacial-resolution respectively, might enable us to build a novel model of human program comprehension processes.

III. Method

A. Subjects

In this study, we recruited ten subjects majoring in computer science from National Institute of Technology, Nara College and Nara Institute of Science and Technology (all males, aged between 19 and 37 years). All subjects had normal or corrected-to-normal vision and understood basic-level Java grammars. One additional subject participated the experiment, but was not included in the analysis because of technical errors on the EEG measurement.

B. Experimental design

Software developers perform diverse decision-makings during program comprehension. In this study, we focused on judgements of conditional expression because the processes are simple enough to be repeated many times and developers frequently face such decision-making processes.

To capture ERPs in programming-related activities, we designed a simple comparison operation task that requires subjects to judge whether a given equation is True or False (Fig.1). For each trial, we presented an equation involving two one-digit integers (0-9) and one operator ($! =, ==, >, >=, <, or <=$) to subjects for 1.5 seconds. A fixation cross was then shown for 1.5 seconds and subjects answered by pressing a button corresponding to “True” or “False”. The notation used in all equations followed the grammar of Java language. Note that we intentionally separated the time slots for conditional expression and response to mitigate potential artefacts caused by motor-related activities especially for finger movements.

The experiment consisted of 10 runs and each single run included 102 trials of the conditional expression tasks (1,020 trials in total for each subject). We made the correct answers of 510 trials be True and the remaining 510 trials be False to balance the numbers of samples between these two cases. Additionally, the number of trials involving each operator was also balanced, yielding 170 trials for each operator. We used PsychoPy (version 1.90.3, [12]) for the stimulus presentations and all behavioral responses recording. The presentation order of each equations was randomized only one time and we applied the same randomized order for all subjects.

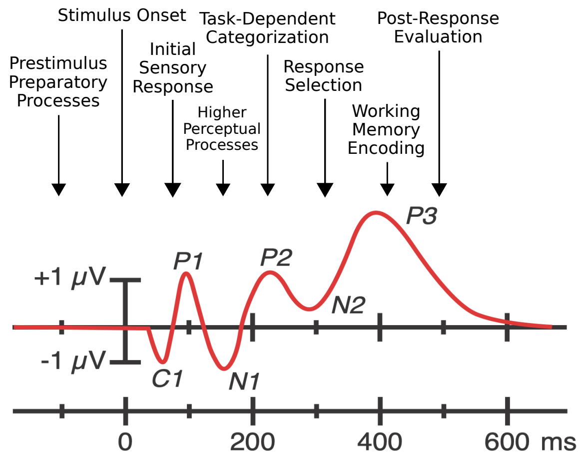


Fig. 2. Example of ERPs and corresponding interpretations. This figure was adopted from the slides of the ERP Boot Camp organized by Steve Luck and Emily Kappenman (https://erpinfo.org/the-erp-boot-camp).

C. Data collection and preprocessing

We used the NeXus-10 MARK II (Nanotech Image Ltd.) to collect EEG data from all subjects while they performed the experimental tasks. We measured subjects' EEG with 256Hz sampling frequency from four different electrode positions; Fpz, Fz, Cz, and Pz based on the 10-20 system [13]; yielding 40,800 data ($1,020 trials \times 10 subjects \times 4 channels$). Responses to each trial and their correctness were also recorded from all subjects as behavioral data.

For preprocessing, we first excluded EEG data measured when the subjects had wrong or no answers and responded too quickly (i.e. in the 'Equation' periods). Through this procedure, 372 data were excluded and 40,428 data remained. We then applied an Infinite Impulse Response (IIR) bandpass filter on the remaining EEG data to remove the task-irreverent artefacts. The low and high cutoff frequencies of the bandpass filter were 2 Hz and 45 Hz, respectively. This preprocessing procedure was performed on EEG data of each trial of each subject independently. We used the preprocessed EEG data and investigated whether their waveforms are consistent with the plausible waveforms of ERPs (Fig.2, [14]).

ERP components are extracted from subject's preprocessed EEG. Each negative peak were labeled as C1, N1, N2, and each positive peak are labeled as P1, P2, P3 according to the visual inspection. If there are multiple peaks within the temporal interval (Table.I), the first point is selected as a peak of the component.

TABLE I
INTERVAL FOR EACH COMPONENTS

Components	Time (msec)
C1	0 – 200
P1	150 – 300
N1	270 – 400
P2	400 – 550
N2	550 – 900
P3	840 – 1100

D. Research Question

We conducted qualitative analyses to address the following research questions:

RQ1 Does ERP appear in EEG data measured during a simple equation judgment task?

RQ2 Are ERP waveforms different between the conditional expression "True" and "False"?

To answer RQ1, we compare the preprocessed EEG data with the plausible waveforms of ERPs shown in Fig.2. Understanding of cognitive process of developer without any interruption should be useful for situation-based developer support on IDE or other software development systems. RQ2 was set to verify whether the ERPs differ depending on the characteristics of the equation. Many studies have shown that ERP waveforms differ in peak amplitudes and peak latencies of their components depending on the type and property of tasks[9], [15]. To answer RQ2, we extract ERP components from waveforms of True condition and False condition.

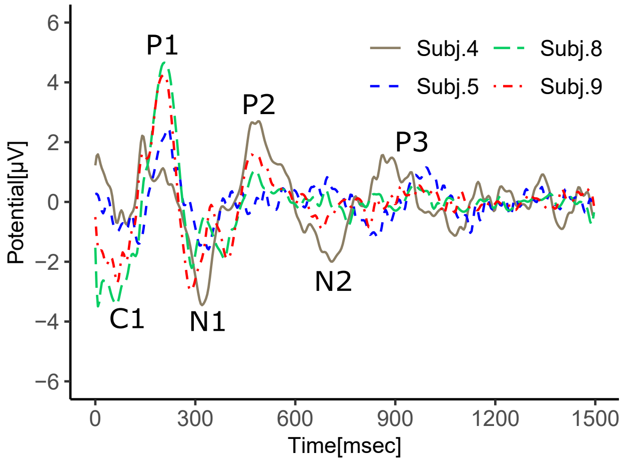
To make the result presentations as simple as possible, this paper only present the results of EEG data measured on the Fpz channel, which showed the most characteristic waveforms among the all four channels.

IV. Results and Discussions

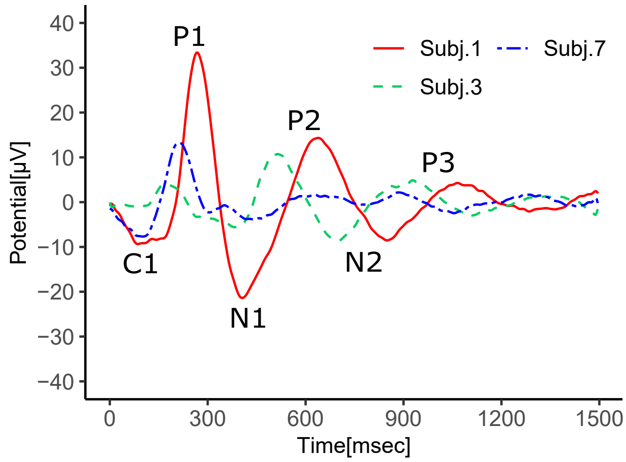
A. RQ1: ERP Detection

Previous studies have demonstrated that ERP waveforms differ between subjects and/or between performing tasks [5], [10]. Here, we first examined the averaged EEG signals of each individual subject to investigate whether ERP waveforms appeared or not during the task. Fig.3 shows averaged EEG of each subject. Four subjects in Fig.3(a) presented clear ERP waveforms that had components around at 100-150 msec (C1), 150-250 msec (P1), 300-350 msec (N1) and so on. Three subjects in Fig.3(b) shows unclear but plausible ERP-like waveforms with large peak amplitude ranging from 40 to -30 μv. However, another three subjects in Fig.3(c) shows different waveforms, potentially contaminated by periodical noises.

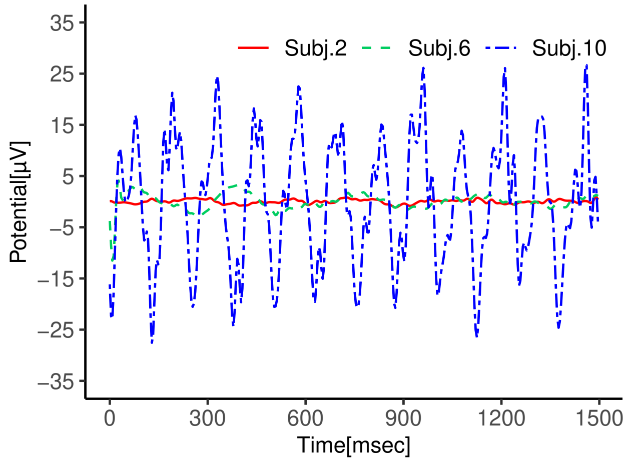
We observed clear ERP waveforms from EEG data of four subjects. Fig.3(a) indicated that the peak latencies of first four components (C1, P1, N1, and P2) were roughly consistent between these subjects. In contrast, the peak latencies and peak amplitudes of the later two components (N2 and P3) seemed to be more diverse than the others. These differences in observed waveforms might be affected by differences between internal cognitive demands. Specifically, the first several components



(a) ERP observed in four subjects. Each wave was obtained by averaging all EEG data based on stimulus-onsets, independently for each subject. The labels (e.g. C1, P1) were provided by the authors.



(b) ERP-like waveforms with large peak amplitudes observed in three subjects.



(c) EEG signals without ERP observed in three subjects.

Fig. 3. Averaged individual EEG for each subject.

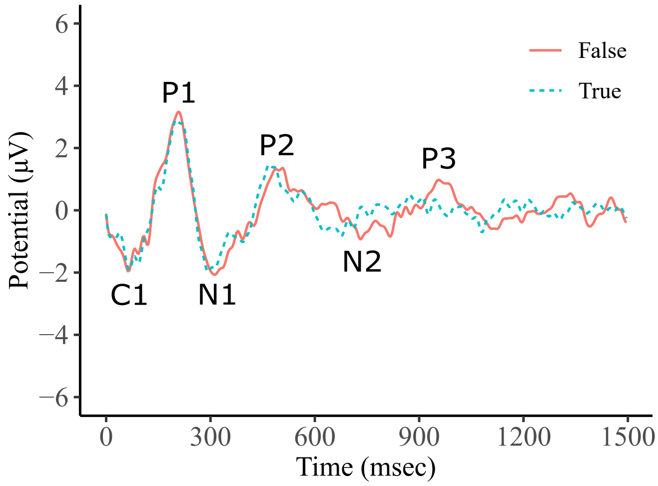


Fig. 4. Averaged waveforms of “True” condition and “False” condition.

may be associated with visual responses while the later peaks might be related to more cognitive demands.

Fig.3(b) shows the averaged signal of other three subjects with larger peak amplitude than those in Fig.3(a). Inter-subject variations in peak amplitude have been observed in numerous EEG studies [16]. Fig.3(c) represents the averaged signals of other three subjects without ERP-like waveforms. The periodic waveform observed in Subj.10 may be affected by periodic artifacts such as head motion. The waves without clear component might indicate that the subjects (Subj.2 and 6) have no significant ERP. From results of individual ERP waveform, our answer to RQ1 is the following.

Answer to RQ1

ERP waveforms appeared in EEG data of four subjects out of ten. Other three subjects showed unclear but plausible ERP-like waveforms.

B. RQ2: Difference of ERP between True and False

In this section, we focus on only four subjects who showed clear ERP waveforms (Subj. 4, 5, 8, and 9) to analyze the difference between conditions, “True” and “False”.

Fig.4 shows averaged ERP waveforms for four subjects of True (called $grandAve_{True}$) and False ($grandAve_{False}$).

The figure shows grand-average ERP waveforms ($grandAve_{True}$ and $grandAve_{False}$) which have components C1, P1, N1, and P2. The peak amplitudes of the components after N2 are a slightly small. In addtion, the difference between $grandAve_{True}$ and $grandAve_{False}$ is not clear. Peak latencies of N2 and P3 of $grandAve_{True}$ are faster than those of $grandAve_{False}$, however the other components (C1, P1, N1, and P2) has almost same peak latencies and peak amplitudes.

One of the reason why the result did not show any different characteristic between $grandAve_{True}$ and $grandAve_{False}$ might be individual differences. Our experiment task demands thinking to subjects for evaluating each conditional expression. Since each subject requires different time to evaluate conditional expression, peak latencies of each components of each subject might appear at different time points. Fig.5 shows individual ERP waveforms for each subjects at conditional expressions that are True (called $indiv_{True}$) and False ($indiv_{False}$). This figure shows that the individual ERP waveforms have clear ERP components, however peak amplitudes and peak latencies differ among subjects. Therefore, we extract peak of each component from the individual ERP waveforms (see III-C) to analyze their tendencies.

Table.II shows the peak amplitudes and latencies of each component. All components of $indiv_{False}$ had larger absolute values of the peak amplitudes and a later peak letencies than those of $indiv_{True}$. Particularly, the peak amplitudes of N1 have a large difference compared with other component. $indiv_{False}$ condition of all subjects have a larger absolute value of the peak amplitude of N1. Although we did not do any statistical analysis due to the lack of the sufficient number of samples, We discuss the two characteristics shown at comparison of $indiv_{True}$ and $indiv_{False}$

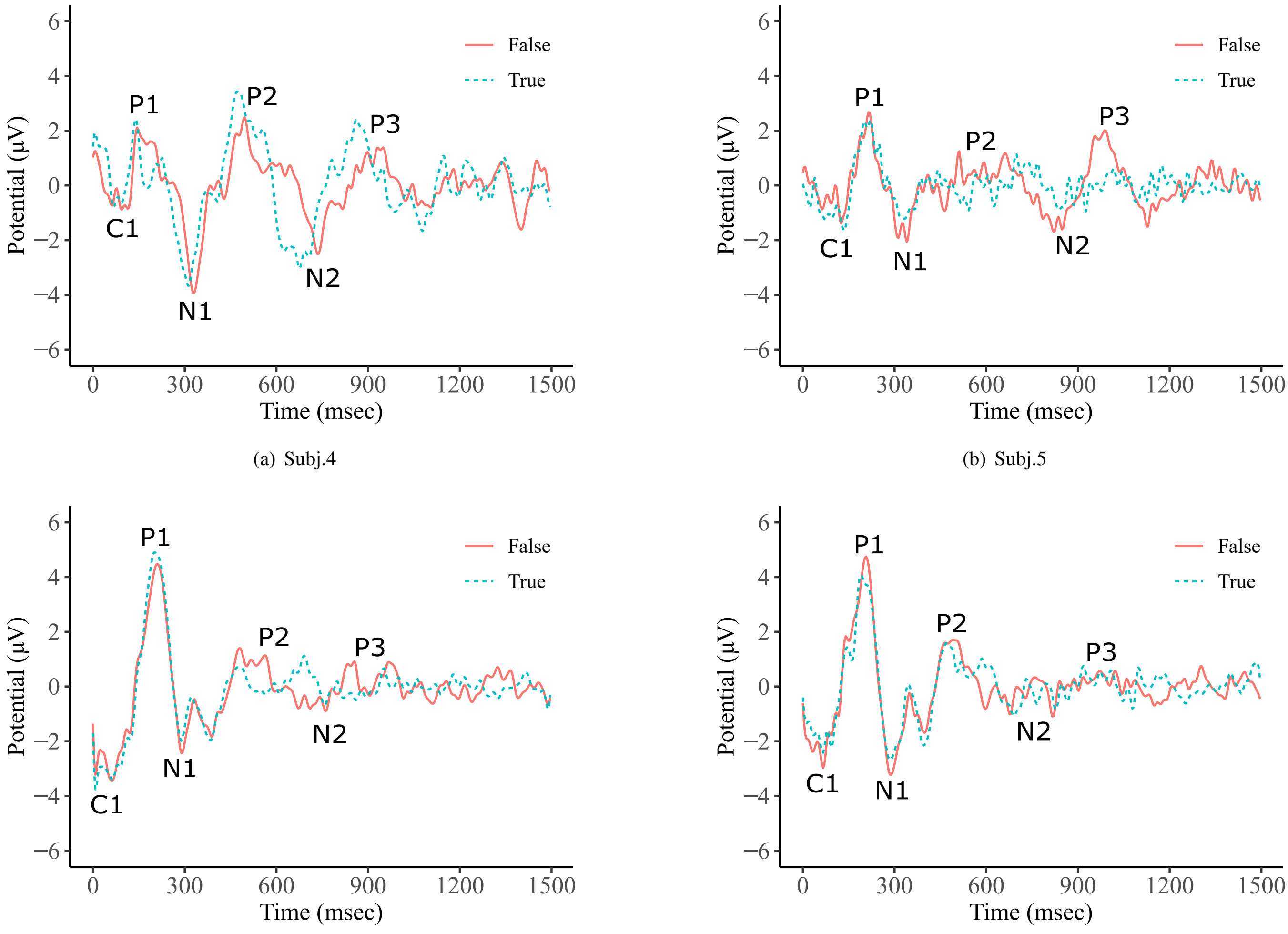
First, the result showed $indiv_{False}$ had the later peak latency than those of $indiv_{True}$. The time difference between $indiv_{True}$ and $indiv_{False}$ components may indicate a difference of task difficulty. In the field of cognitive psychology, it has been well known that performing two tasks at the same time decreases task performance (dual-task interference). Kimura and Takeda[17] reported that task difficulty of visual stimuli affected the peak latency of an ERP component (called DRM). The interference in the dual-task increases subject’s mental work load (MWL), and it can exceed the capacity limit of the subject. Although the tasks given to the subjects in this study were not dual-task, it is possible that there is a difference in the effect of $indiv_{True}$ and $indiv_{False}$ toward subjects’ MWL, and appeared as the difference in peak latency.

Second, our result showed that $indiv_{False}$ had the larger absolute peak amplitude than $indiv_{True}$ in every components, especially in N1. N1 amplitude is a parameter that has commonly been linked to early attentional selection processes and allocation of perceptual resources [18], [19], [20]. Solís-Marcos and Kircher measured ERPs during single, dual, and triple tasks which leads to different MWL to participants. The result showed that N1 amplitude at single task was lower than dual and triple tasks [18]. In our study, it could be possible that the difference of conditional expressions affect subjects’ MWL differently, and it lead to the difference of peak amplitude and peak latency.

From all results based on the comparison of $indiv_{True}$ and $indiv_{False}$, our answer to RQ2 is the following.

Answer to RQ2

The ERP waveforms are different between True “conditional expression” are and False one. On average, the absolute value of the peak amplitude of $indiv_{False}$ is larger than that of $indiv_{True}$, and the peak latency of $indiv_{False}$ is later than that of $indiv_{True}$.



(a) Subj.4 (b) Subj.5 (c) Subj.8 (d) Subj.9

Fig. 5. True and False ERPs of each subject.

TABLE II
PEAK AMPLITUDE AND PEAK LETENCY OF EACH COMPONENT

Type		Peak amplitude (μV) C1	P1	N1	P2	N2	P3	Peak latency (msec) C1	P1	N1	P2	N2	P3
Average	False	-2.17	3.48	-2.91	1.71	-1.55	1.22	90.8	199.2	315.5	499.0	787.1	945.3
	True	-2.15	3.30	-2.39	1.57	-1.42	1.10	72.3	195.3	308.6	468.8	744.1	940.4
SD	False	1.06	1.16	0.72	0.48	0.63	0.54	25.1	27.2	22.8	11.2	36.5	51.5
	True	1.09	1.23	0.90	1.14	0.94	0.75	44.2	27.2	18.9	13.8	63.2	56.6

V. CONCLUSIONS

To investigate EEG data with more fine-grained temporal resolution, we challenged to assess the feasibility of ERP analyses in the software engineering domain. We conducted an experiment to capture ERPs of ten subjects during a simple program comprehension process. As a result, ERP waveforms were detected in seven subjects out of ten; four showed clear and three showed plausible ERP-like waveforms. We also found that ERP waveforms of "True condition" and "False condition" have different shape to each other. Although we need more detailed analysis based on the rich knowledge in neuroscience and cognitive science, the result of this study suggests ERP analysis would be beneficial for software engineering study to reveal detail of program comprehension processes in human brain.

As a future work, it is necessary to discuss more why the peak amplitude of $indiv_{False}$ is higher and the peak latency is later than that of $indiv_{True}$. Statistical analysis with larger number of participants are also required for reliable results. It is also necessary to consider the ERP extraction methods and component selection. In general, the early ERP components reflect basic sensory processing of stimuli at a lower level of processing, and the later ERP components reflect the perceptual and cognitive processing of stimuli at a higher level of processing [21]. The analysis of later ERPs is an interesting research target due to the characteristics of program comprehension. Therefore, construction of experimental design for measuring later ERPs such as P3 (third positive peak) during code reading is one of important future work.

VI. Acknowledgments

This work was supported by JSPS KAKENHI Grant Number JP18J22957 and JP16K00114.

References

[1] T. Dyba, B. A. Kitchenham, and M. Jorgensen, "Evidence-based software engineering for practitioners," *IEEE software*, vol. 22, no. 1, pp. 58–65, 2005.

[2] T. Roehm, R. Tiarks, R. Koschke, and W. Maalej, "How do professional developers comprehend software?" in *Proceedings of the 34th International Conference on Software Engineering*. IEEE Press, 2012, pp. 255–265.

[3] T. Fritz, A. Begel, S. C. Müller, S. Yigit-Elliott, and M. Züger, "Using psycho-physiological measures to assess task difficulty in software development," in *Proceedings of the 36th international conference on software engineering*. ACM, 2014, pp. 402–413.

[4] S. C. Müller and T. Fritz, "Stuck and frustrated or in flow and happy: sensing developers' emotions and progress," in *2015 IEEE/ACM 37th IEEE International Conference on Software Engineering*, vol. 1. IEEE, 2015, pp. 688–699.

[5] T. Picton, S. Bentin, P. Berg, E. Donchin, S. Hillyard, R. Johnson, G. Miller, W. Ritter, D. Ruchkin, M. Rugg *et al.*, "Guidelines for using human event-related potentials to study cognition: recording standards and publication criteria," *Psychophysiology*, vol. 37, no. 2, pp. 127–152, 2000.

[6] X. Xia, L. Bao, D. Lo, Z. Xing, A. E. Hassan, and S. Li, "Measuring program comprehension: A large-scale field study with professionals," *IEEE Transactions on Software Engineering*, vol. 44, no. 10, pp. 951–976, 2017.

[7] J. Siegmund, C. Kästner, S. Apel, C. Parnin, A. Bethmann, T. Leich, G. Saake, and A. Brechmann, "Understanding understanding source code with functional magnetic resonance imaging," in *Publication:ICSE 2014: Proceedings of the 36th International Conference on Software Engineering*.

[8] S. Lee, A. Matteson, D. Hooshyar, S. Kim, J. Jung, G. Nam, and H. Lim, "Comparing programming language comprehension between novice and expert programmers using eeg analysis," in *2016 IEEE 16th International Conference on Bioinformatics and Bioengineering (BIBE)*. IEEE, 2016, pp. 350–355.

[9] M. Kutas and S. A. Hillyard, "Event-related brain potentials to grammatical errors and semantic anomalies," *Memory & cognition*, vol. 11, no. 5, pp. 539–550, 1983.

[10] M. Kutas and K. D. Federmeier, "Thirty years and counting: finding meaning in the n400 component of the event-related brain potential (erp)," *Annual review of psychology*, vol. 62, pp. 621–647, 2011.

[11] E. F. Lau, C. Phillips, and D. Poeppel, "A cortical network for semantics:(de) constructing the n400," *Nature Reviews Neuroscience*, vol. 9, no. 12, p. 920, 2008.

[12] J. W. Peirce, "Psychopy-psychophysics software in python," *Journal of neuroscience methods*, vol. 162, no. 1-2, pp. 8–13, 2007.

[13] H. H. Jasper, "The ten-twenty electrode system of the international federation," *Electroencephalogr. Clin. Neurophysiol.*, vol. 10, pp. 370–375, 1958.

[14] S. Luck and E. Kappenman, "UC-Davis / SDSU ERP boot camp," https://erpinfo.org/the-erp-boot-camp.

[15] M. A. Pitts, J. L. Nerger, and T. J. R. Davis, "Electrophysiological correlates of perceptual reversals for three different types of multistable images," *Journal of Vision*, vol. 7, no. 6.

[16] R. van Dinteren, M. Arns, M. L. A. Jongsma, and R. P. C. Kessels, "P300 development across the lifespan: A systematicreview and meta-analys," *POLS ONE*, vol. 9, no. 2, 2014.

[17] M. Kimura and Y. Takeda, "Task difficulty affects the predictive process indexed by visual mismatch negativity," *Frontiers in human neuroscience*, vol. 7, no. 267, 2013.

[18] I. Solís-Marcos and K. Kircher, "Event-related potentials as indices of mental workload while using an in-vehicle information system. cognition," *Cognition, Technology & Work*, vol. 21, no. 1, pp. 55–67, 2019.

[19] A. F. Kramer, C. D. Wickens, and E. Donchin, "An analysis of the processing requirements of a complex perceptual-motor task," *Human factors*, vol. 25, no. 6, pp. 597–621, 1983.

[20] A. Kok, "Event-related-potential (erp) reflections of mental resources: a review and synthesis," *Biological psychology*, vol. 45, no. 1-3, pp. 19–56, 1997.

[21] C. Portella, S. Machado, F. Paes, M. Cagy, A. T. Sack, A. Sandoval-Carrillo, J. Salas-Pacheco, A. C. Silva, R. Piedade, P. Ribeiro, A. E. Nardi, and O. Arias-Carrión, "Differences in early and late stages of information processing between slow versus fast participants," *International archives of medicine*, vol. 7, no. 49, 2014.