Brain Activity Measurement during Program Comprehension with NIRS

Yoshiharu Ikutani
Department of Information Engineering
Nara National College of Technology
Japan, Nara, Yamatokoriyama
Email: ikutani@info.nara-k.ac.jp

Hidetake Uwano
Department of Information Engineering
Nara National College of Technology
Japan, Nara, Yamatokoriyama
Email: uwano@info.nara-k.ac.jp

***Abstract*—Near infrared spectroscopy (NIRS) has been used as a low cost, noninvasive method to measure brain activity. In this paper, we experiment to measure the effects of variables and controls in a source code to the brain activity in program comprehension. The measurement results are evaluated after noise reduction and normalization to statistical analysis. As the result of the experiment, significant differences in brain activity were observed at a task that requires memorizing variables to understand a code snippet. On the other hand, no significant differences between different levels of mental arithmetic tasks were observed. We conclude that the frontal pole reflects workload to short-term memory caused by variables without affected from calculation.**

I. Introduction

Near infrared spectroscopy (NIRS) has been used in various research fields as a low cost, noninvasive method to measure brain activity. Using the differential of light absorptivity between oxy-hemoglobin(Oxy-Hb) and deoxy-hemoglobin(deOxy-Hb), NIRS measures a cerebral blood flow to estimate a regional brain activity [1].

In the domain of program comprehension, Siegmunt et al. insisted on the need to identify the brain areas that are activated during program comprehension [2]. To quantify an action on program comprehension, Nakagawa et al. measured cerebral blood flow of participants who simulate source code psychologically using NIRS[3]. NIRS is an especially suitable method for a program comprehension research in various devices to measure brain activity, because of high temporal resolution and low restriction on participant. Program comprehension may consist of many factors such as number calculation, variable memorizing, and understanding of conditional branches. However, effects of each factors in program comprehension to brain activity is unclear. To understand how programmers comprehend program source code, observation of each factor's effect is required. Brain activity measurement in program comprehension proceeds a quantitative analysis of program comprehension based on Neuroscience and Neuropsychology.

In this paper, we investigate the effects of each factors to the brain activity. We focus on the two factors in the experiment; variables and control statements. Variables and control statements are primary elements in source code, and important factors in program comprehension. We measure the participant's brain activity during tasks that read the code snippet, and compare them statistically.

II. Related work

A. Brain Measurement with NIRS

NIRS is a low cost and non-invasive method to measure brain activity. Various research fields use the device to measure brain activities that related to language, auditory, motor functions[4]. NIRS is light weight and tolerant towards electrical noises, hence, the device is applied to BMI(Brain Machine Interface) domain as a method to measure brain activity[5].

NIRS has higher temporal resolution and lower restriction on participant in comparison with other methods such as PET, fMRI, EEG, MEG[6]. On the other hand, spatial resolution of the method is low and the measure depth is restricted to a surface of the brain. Also the most of NIRS devices measure brain activity as a relative value based from the initial value. Therefore comparison of measurement values on different conditions is difficult. Additionally, the method is sensitive to physiological noises that include body motion, swing of transmit cable, heartbeat, and respiration[7]. Therefore, proper noise reduction is required to analyze the experiment result.

To cope with the such problems, several methods have been proposed. Mitsuya et al. focused on time series of signals obtained from NIRS, and proposed a method to eliminate biological noises using trend analysis and moving average method[8]. Tsunashima et al. proposed a noise reduction/normalization method using discrete wavelet transform and Z-score transformation to statistical processing[9]. In this study, we use the Tsunashima's method for noise reduction and statistical analysis.

B. Brain Measurement in Program Comprehension

Siegmunt et al. stated the need to identify the brain areas that are activated during program comprehension, and proposed an experimental design to measure program comprehension based on fMRI[2]. An introduction of brain activity measurement into program comprehension research allows us to observe what is happening inside the brain during program comprehension directly. The results of the measurement are essential information to understand the difference between good programmer and bad programmer, or to develop the programmer support systems.

Only a few studies report the measurement result of brain activity during program comprehension. Nakagawa et al.

measured brain activity during program comprehension using NIRS[3]. The study measured participant's cerebral blood flow during simulating a source code to quantify program comprehension process. The result of the experiment showed that brain activities differed according to the task difficulty, and the largest positive blood flow was observed at from initial to middle phase of the task.

Program comprehension consists of many factors such as number calculation, variables memorizing, and understanding of conditional branches. That is, the code simulation task consist of several factors that affect differently to the brain activity. To understand the relationship between brain activity and program comprehension, measuring the effect of each factors on brain activity is essential. In this study, we measure the effects of variable memorizing and conditional branches through an experiment.

III. Measurement with NIRS

A. NIRS

Increasing of neural activity is accompanied by blood flow increasing. In order to estimate brain activity, NIRS measures blood flow change in the brain that follows neural activity. Compare with the other methods, NIRS has some advantages such as high temporal resolution and low restriction of participants. The high temporal resolution allows us to analyze brain activity in detail, and the low restriction enables measurement on more practical conditions.

An increase in blood flow causes haemoglobin density changes in the same region. Specifically, Oxy-Hb increase and deOxy-Hb decrease are observed in a region where neural activity increases[10]. Oxy-Hb and deOxy-Hb have different absorptivity to near infra-red light. NIRS measures blood flow changes related to neuronal activity by observe the difference of near infra-red light[4]. In general, Oxy-Hb is considered as the best index of brain measurement experiment with NIRS. However most of NIRS devices that are now broadly used cannot measure optical path length that is essential for identifying blood flow changes as an absolute value. Consequently, measuring result becomes a relative value derived from the value that is measured at start-up. It means that comparison between individuals or regions in the brain is inappropriate. Also NIRS is sensitive to physiological noises such as body motion, heartbeat, and respiration[7]. Therefore, proper noise reduction is required to analyze the experiment result.

B. Wavelet-based multi resolution analysis

Fig.1 shows a measurement result in our experiment using NIRS. The horizontal axis represents a time course during a task, and the vertical axis represents an Oxy-Hb value; the higher value means the higher regional brain activity. The figure shows that the wave contains many noises derived from heartbeats and/or respirations. In this study, we use an wavelet-based method proposed by Tsunashima et al. for noise reduction[9].

Discrete wavelet transform (DWT) is a method that decomposes signals into an approximated component and detailed components. Fig.2 shows the result of decomposing the wave described in fig.1. In the figure, $d1$-$d15$ mean the decomposed components of the original wave and each has different frequency. The components that higher than 1Hz (from $d1$ to $d7$) are considered as measurement noises caused by body motion. Also the 0.015-0.50Hz component ($d8$ and $d9$) and the 0.005-0.15Hz component ($d10$ and $d11$) are noises caused by respiration or blood pressure change. In contrast, low frequency components ($d12$, $d13$, and $d14$) are probably signals derived from brain activity that we focused on the experiment. A noise-reduced wave is reconstructed by removing the noise components and combining the residual components. Procedure of the noise reduction method that we use in this research is described as follows [9]. First, decompose an wave into some components that each has different frequency. The components decided as a noise are removed, then noise-reduced wave is reconstructed from the residual components.

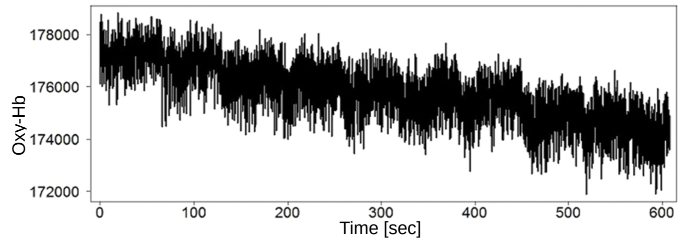


Fig. 1. Original signal from NIRS

C. Z-score

Collected signals from NIRS are quantity of relative changes using the start-up value as a reference: hence comparison of the signals between subjects is inadequate. In this research, we normalize signals into Z-scores to compare the signals between subjects and to adopts a statistical analysis [9]. The method convert the NIRS signals reconstructed by the procedure described in section III-B into Z-scores using the following expression. Here, X is noise-reduced signal during one task, μ and σ are a mean value and a standard deviation respectively.

$$Z = \frac{X - \mu}{\sigma} \qquad (1)$$

IV. Experiment

Two type of tasks were prepared for the experiment, and participant's brain activity during the tasks were measured using NIRS. Eleven male undergraduate students participated in the experiment. All were right-handed and finished their first programming lecture before the experiment.

A. Task

Two types of tasks were prepared in this experiment. In $Program$ task, participants read a code snippet to calculate the value of variables. Participants on $Arithmetic$ task answer a mental arithmetic question. Both tasks consist from three problem sessions and four rest sessions. At before each problem session, a rest session that is used as a baseline of the subsequent problem session is performed. Both problem and rest session continue 32 seconds. During the rest session,

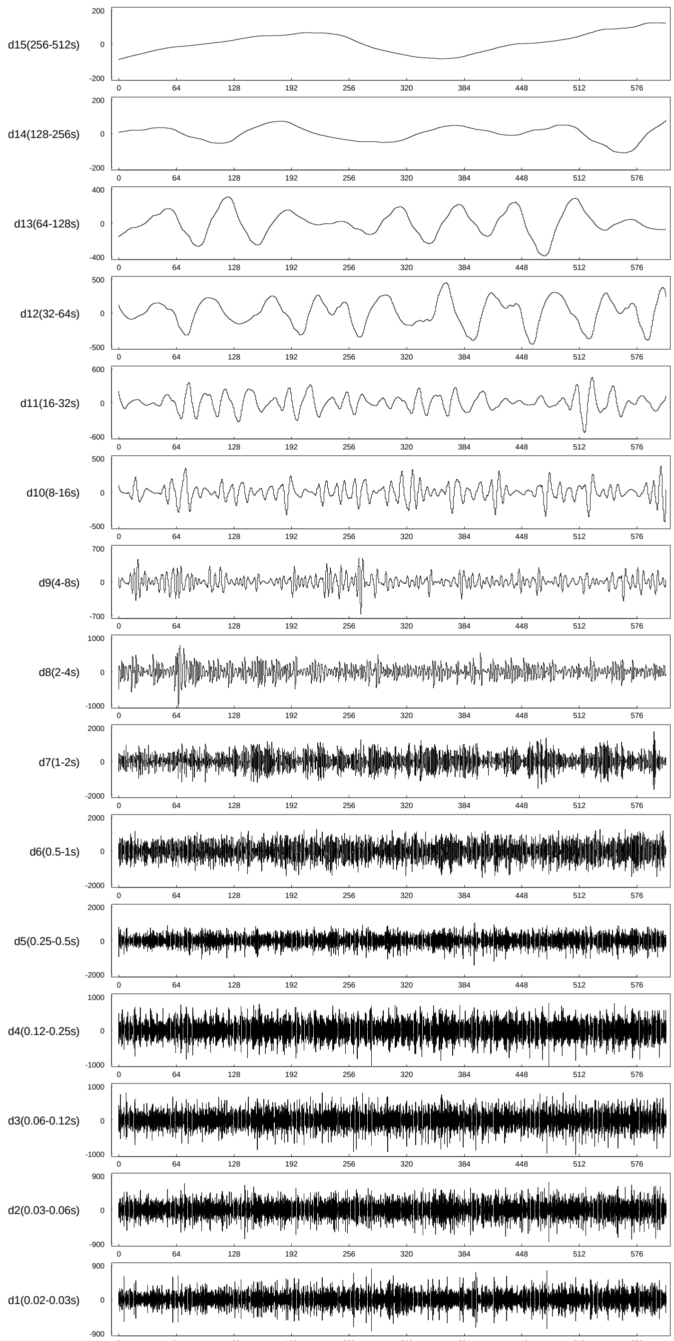


Fig. 2. Decomposed waves

subjects are asked to gaze a cross marker displayed on a center of a computer screen.

1) Program Task: Subjects are asked to answer a consecutive code snippet question displayed on a screen silently. Three types of question are used. All types of questions require to calculate three variables (a, b, and c) in their mind. Fig.3 shows an example of $Program$ task. Details of each questions are as follows:

- Numeric
  Each question consists of three lines of code. Each line calculates a value of variable from three integers.

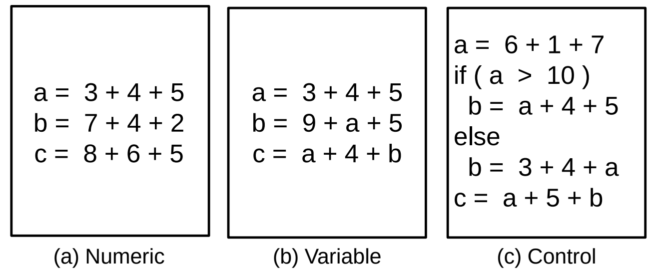


Fig. 3. Example of Program Task

- Variable
  Each question consists of three lines of code. Each line calculates a value of variable from integers and other variables. This question requires to memorize the value of the variables.
- Control
  Each question consists of six lines of code include the if – else statement. Subjects read three lines from the code snippet according to the condition. This question requires memorizing the value of the variables and judge the if – else branch.

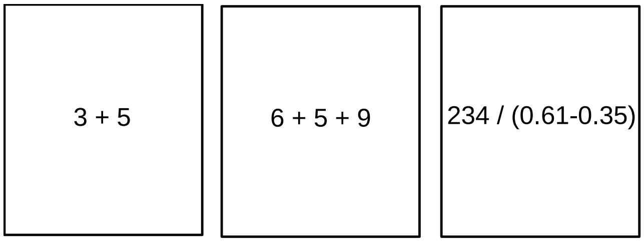
A number of questions in each session is adjusted according to workload of each question. In the experiment, we assign four questions at $Numeric$ session (8 seconds per question), three questions at $Variable$ session (10.6 seconds per question), and two questions at $Control$ session (16 seconds per question) respectively.

While $Numeric$ consists of only numbers, $Variable$ consists of numbers and variables. We expect that an effect of a variable on the brain activity is observed between $Numeric$ and $variable$. Similarly, $Control$ consists of same factors of $Variable$ and a control statement (if-else branch). We therefore expect that an effect of a control statement is observed between $Variable$ and $Control$.

2) Arithmetic Task: Subjects are asked to answer a consecutive of mental arithmetic question displayed on a screen silently. This task is the same task of Tsunashima's experiment [9]. We employed the task to validate the our experimental setting. Three difficulty levels of question are used in the task. All types of questions require to calculate an answer of an equation. Fig.4 shows an example of $Arithmetic$ task. Details of each questions are as follows:

- Low
  Addition of two one-digit numbers.
- Middle
  Addition of three one-digit numbers.
- High
  Subtraction and division of two decimals and one three-digit number.

A number of questions in each session is adjusted according to workload of each question. In the experiment, we assign 16 questions at Low session (2 seconds per question), 10 questions at $Middle$ (3.2 seconds per question), and two questions at $High$ (16 seconds per question) respectively.



(a) Low (b) Middle (c) High

Fig. 4. Example of Arithmetic Task

(a) NeXus10 (b) Measurement state

Fig. 5. Measurement device

The experiment follows the procedure below.

1) Explain the experiment
2) Set the NIRS device to the participant
3) Perform $Program$ task and recode brain activity
4) Three minutes break
5) Perform $Arithmetic$ task and recode brain activity
6) Remove the device

B. Environment

One-ch NIRS device NeXus10 (TMS international BV) is employed in the experiment to record a brain activity. Fig.5 shows the appearance of the device and measurement state. The experiment is performed in a silent room which one subject and two observers are remain. In order to restrain artifacts caused by subject's body motion, he sit on a chair which has armrests and a headrest, and are asked to be on steady condition.

We set the device on the forehead of subject, and measure his brain activity at a sampling frequency 128[Hz]. The measured region is the front of the frontal lobe, i.e. "frontal pole". Frontal pole is considered that it relates to short-term memory and higher-order function like planning an action. It is expected that the region is activated in $Program$ task by memorize the value of variables and by judging the if–else conditions.

C. Data process

The NIRS device measures the changes in Oxy-Hb, deOxy-Hb, and Total-Hb that means the sum of the changes in Oxy-Hb and deOxy-Hb. We use the Oxy-Hb as an evaluation metric for brain activity because the value has a better reflectivity to blood flow changes than the other metrics[10].

First, the original Oxy-Hb signal is decomposed into several components that each has different frequency. A cerebral blood flow reflects a neural activity slowly at a second-order. Therefore the components that have a frequency higher than 1Hz are probably measurement noises caused by subject's body motion and/or others. Also the 0.015-0.50Hz components are noises caused by respiration, and the 0.005-0.15Hz components are noises caused by blood pressure change[7]. The components regarded as noises are eliminated, and then the residual components ($d12$, $d13$, and $d14$ in Fig.2) are reconstructed as a task-related signal. Then, converts the noise-reduced signal into Z-score to enable comparison between subjects and statistical analysis. Finally, calculates brain activity that changes by each session of the task. The brain activity that originated from each problem session is calculated from the difference between the problem session and the former rest session. Let f_0 is a Oxy-Hb value of a session start and f_n is a value of a session end, the brain activity during the problem session $F(t)$ is denoted as following equation.

$$F(t) = \{f_0, f_1, f_2, \cdots, f_{n-1}, f_n\}$$

Also, the brain activity during the rest session which before the $F(t)$ is denoted as following equation.

$$R(t) = \{r_0, r_1, r_2, \cdots, r_{n-1}, r_n\}$$

The brain activity that originated from each problem session $activity$ is formulated as follow:

$$\begin{aligned} rest &= \frac{1}{n}\sum_{i=0}^{n} R(t) \\ activity &= \sum_{i=0}^{n} (f_i - rest) \cdot \tau \end{aligned}$$

Here, $rest$ means the average value of the former rest session, and τ denotes the inverse of sampling frequency.

V. Result

A. Program Task

Fig.6 shows the averaged brain activity of all subjects in $Program$ task. The horizontal axis represents a task time, the vertical axis represents an Oxy-Hb value converted into Z-score. Each character on the top shows the session type at each period. The figure describes that the brain activity increased at the problem sessions and decreased at the rest sessions. Also the figure shows that the brain activity was much increased at $Variable$ than the other problem sessions.

Fig.7 shows the brain activity of all subjects in each problem session at $Program$ task. The vertical axis shows the average $activity$; higher value means higher brain activity. The figure shows the brain activity at $Variable$ session is higher than the other sessions. The result of the Ryan's method describes significant differences between $Numeric$ – $Variable$ (p=0.01) and $Variable$ – $Control$ (p=0.003). In the $Variable$ session, subjects were required to memorize

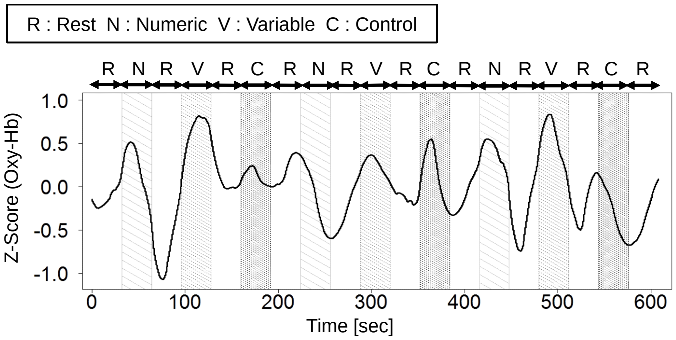


Fig. 6. Brain activity in $Program$ task

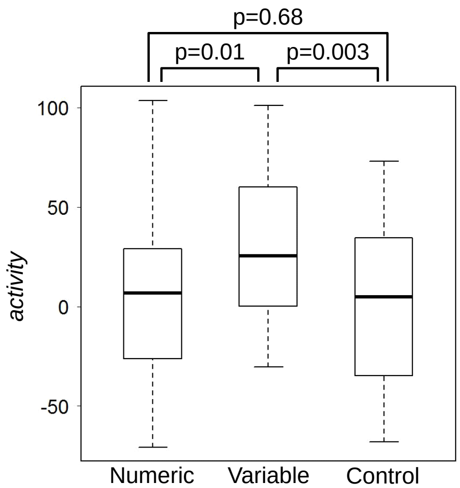


Fig. 7. Brain activity of each problem session in $Program$ task

the value of variables. The result suggests a workload to short-term memory caused by memorizing variables increases brain activity. The result indicates that frontal lobe relates to understand variables in a source code.

In contrast with $Variable$, $Control$ shows no differences between other problem sessions. The possible causes of the result are follows:

- processing the condition of if–else statement does not affect brain activity of frontal lobe.
- the fewer number of problems on $Control$ session (two problems per session) compared with $Variable$ session (three problems per session) reduces the workload.

We assigned the least number of the questions in $Control$ session, because each question of $Control$ session has the largest code snippet. However the number of lines which subjects actually read during the $Control$ question (four lines) is similar to other questions (three lines in $Arithmetic$ and $Variable$), hence no difference were observed in the experiment. Clarifying the effect of if–else and other control statement to the brain activity is our future work.

B. Arithmetic Task

Fig.8 shows the averaged brain activity of all subjects in $Arithmetic$ task. The figure shows a large increase in

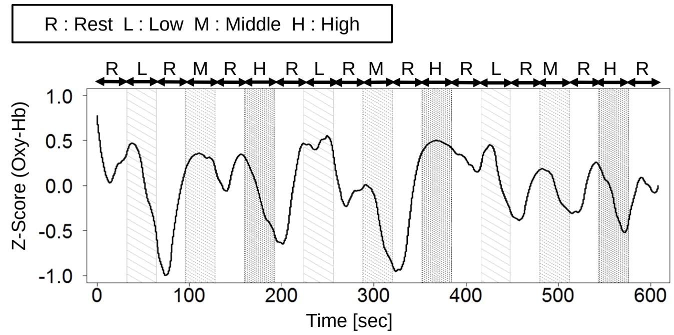


Fig. 8. Brain activity in $Arithmetic$ task

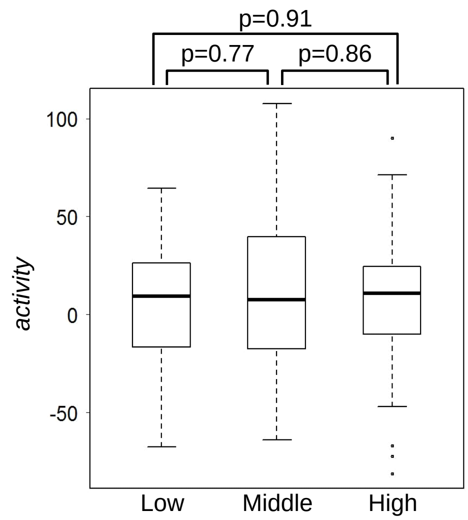


Fig. 9. Brain activity of each problem session in $Arithmetic$ task

brain activity at the first $Middle$ session. On the other hand, although the second session of Low and $High$ increased largely, the brain activity in the second $Middle$ session was decreased. Fig.9 shows the brain activity of all subjects in each problem session at $Arithmetic$ task. There were no significant differences between sessions. Tsunashima et al. showed that there were significant differences between Low – $High$, and $Middle$ – $High$ of the $Arithmetic$ task [9]. However there was no significant difference between the difficulty in our experiment. In our experiment, we select the front of the frontal lobe (frontal pole) as a measurement region, and Tsunashima et al. measured the Dorsolateral left prefrontal cortex. Frontal pole is considered that it relates to short-term memory and high level cognitive activity like planning an action. The measurement result with fMRI showed that no significant difference appeared between the tasks on the frontal pole [9]. These results indicate that the frontal pole is not activated by calculation in contrast to Program Task, and that activated brain regions differ by type of tasks; therefore, selection of target region at measurement of the program comprehension is important.

VI. Discussion

A. Brain Activity in Program Comprehension

The result of the $Program$ task describes brain activity on $Variable$ was higher than $Numeric$ and $Control$. In the

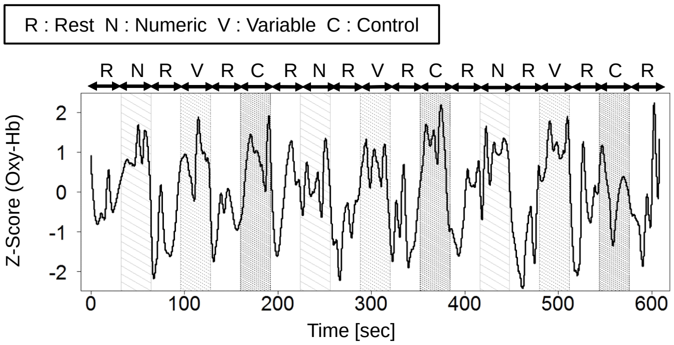


Fig. 10. Reconstruct wave include d10 and d11 component

$Program$ task, variable memorizing and calculation of the variables are the possible cause of the brain activity. However, the result of the $Arithmetic$ task shows that the workload from the calculation is not affected to the frontal pole. Hence, the experiment result shows that a frontal pole reflects workload to short-term memory caused by variables without affected from calculation. Most of program include simple calculation such as increment of index value. Measurement of programmer's frontal pole may allows us to analyze the effect of memorization or judgment in program comprehension.

B. Time Resolution of Brain Activity

The noises in the original signals were eliminated by wavelet-based multi resolution analysis in our experiment. Because of the wavelet transform characteristic, low-frequency components have a coarse time resolution, therefore, the time resolution of reconstructed waves is becomes coarse. This noise reduction method therefore reduces one of the advantages of NIRS: high sampling frequency. More fine grained time resolution analysis with noise-reduction will enable to analyze brain activity changes on program comprehension in detail.

In this experiment, we eliminated the components regarded as noises (from $d1$ to $d11$ and $d15$ in Fig.2) and reconstructed the task-related signal from the residual signals ($d12$, $d13$, and $d14$). A cerebral blood flow reflects a neural activity slowly at a seconds-order. Hence, the components that has frequencies higher than 1Hz (from $d1$ to $d6$) are considered as measurement noises. However, the components which cycle ranged from 8 seconds to 32 seconds ($d10$ and $d11$) may contain brain activity changes caused by tasks. Fig.10 shows the reconstructed wave from $d10$ to $d14$ component. In this figure, brain activity in $Numeric$ session (four questions per session) moves up and down more frequently compared with $Control$ session (two questions per session). The figure suggests the proper component selection for analysis target task is required.

VII. Conclusion

In this paper, we measured the brain activity on two task types to investigate the effects of variables and controls in a source code during program comprehension. We measured blood flow of frontal pole during the tasks that read the code snippet or mental arithmetic problem. Noise reduction using wavelet-based multi resolution analysis and normalization by Z-score conversion were used for statistical comparison between the problem sessions in each task.

As a result, significant differences in brain activity between $Variable$ and other problem sessions in $Program$ task were observed. In contrast, no significant difference was observed in $Arithmetic$ task; that means workload that derives from calculation has no effect on the frontal pole. The result suggests that the frontal pole reflects workload to short-term memory caused by variables without affected from calculation. Therefore, measuring the frontal pole activity is an useful procedure to quantify the workload on short-term memory during the program comprehension task.

In this paper, the effect of the if–else statement to brain activity is not observed. To clarify the effect of control statement, additional experiments that adjusts the number of the question at $Variable$ and $Control$ session is required. As a future work, we plan to brain activity analysis of other control statement such as ”for” and/or ”while”. Also the measurement of large size source code such as a function or whole program is a interesting setting.

References

[1] E. Watanabe, Y. Yamashita, A. Maki, Y. Ito, H. Koizumi, “Non-invasive Functional Mapping with Multi-channel Near Infra-red Spectroscopic Topography in Humans,” Neuroscience Letters, Vol.205, pp.41–44, 1996.

[2] J. Siegmund, A. Brechmann, S. Apel, C. Kastner, J. Liebig, T. Leich, G. Saake, “Toward Measuring Program Comprehension with Functional Magnetic Resonance Imaging,” In Proceedings of the ACM SIGSOFT 20th International Symposium on the Foundations of Software Engineering (FSE ’12), No.24, 2012.

[3] T. Nakagawa, Y. Kamei, H. Uwano, A. Monden, K. Matsumoto, “Quantifying Program Understanding Activities Based on Cerebral Blood Flow Measurement,” In proceedings of 20th Workshop on Foundation of Software Engineering (FOSE2013), pp.191–196, 2013(in Japanese.)

[4] M. Ferrari, V. Quaresima, “A Brief Review on the History of Human Functional Near-Infrared Spectroscopy(fNIRS) Development and Fields of Application,” Neuroimage, Vol.63, No.2, pp.921–35, 2012.

[5] S. Coyle, T. Ward, C. Markham, G. McDarby, “On the Suitability of Near-Infrared(NIR) Systems for Next-Generation Brain-Computer Interfaces”, Physiological Measurement, Vol.25, pp.815–822, 2004.

[6] G. Strangman, J. P. Culver, J. H. Thompson, D. A. Boas, “A Quantitative Comparison of Simultaneous BOLD fMRI and NIRS Recordings during Functional Brain Activation,” Neuroimage, Vol.17, No.2, pp.719–731, 2002.

[7] S. Coyle, T. Ward, C. Markham, “Physiological Noise in Near-infrared Spectroscopy: Implications for Optical Brain Computer Interfacing”, In Proceedings of the 26th Annual International Conference on Engineering in Medicine and Biology, pp.4540–4543, 2004.

[8] R. Mitsuya, H. Morikawa, S. Kurihara, T. Kawai, Y. Tamura, “Removal of Organism Noise in Near-Infrared Spectroscopy Signal by Time Series Analysis,” The transaction of Human Interface Society, Vol.11, No.2, pp163–172, 2009(in Japanese.)

[9] H. Tsunashima, K. Yanagisawa, M. Iwadate, “Measurement of Brain Function Using Near-Infrared Spectroscopy (NIRS)”, Neuroimaging-Methods, Prof. Peter Bright (Ed.), 2012.

[10] Y. Hoshi, N. Kobayashi, M. Tamura, “Interpretation of Near-Infrared Spectroscopy Signals: A Study with A Newly Developed Perfused Rat Brain Model,” Journal of Applied Physiolosy, Vol.90, No.5, pp.1657–1662, 2001.