

脳血流計測に基づくプログラム理解行動の定量化

Quantifying Program Understanding Activities Based on Cerebral Blood Flow Measurement

中川 尊雄* 亀井 靖高† 上野 秀剛‡ 門田 暁人§ 松本 健一**

あらまし 本論文では、プログラム理解中の人物が脳をどの程度使ってプログラムを読んでいるかや、対象の理解難度を定量化するため、脳血流計測を用いた被験者実験を行った。実験の結果、(1) プログラムの読解難度が高い（難読化されている）ほど脳活動が活発になる、(2) 特に活発化するの読解の序盤から中盤にかけて、であることがわかった。この結果は、プログラム理解中の人物がどの程度プログラム理解に苦勞しているかや、プログラムの理解難度を脳血流計測によって定量的に評価できる可能性を示唆している。

1 はじめに

プログラム理解は、情報の記憶や想起、認知、注意など、内的な認知プロセスを含む知的活動であり、一般的にその直接計測は困難とされる。神経科学の分野では、認知プロセスの計測を目的とし、脳活動を計測するアプローチが広く用いられている[1]。

Siegmunt らは、こうした神経科学的アプローチを応用し、プログラム理解と関係する脳部位を同定するため、脳血流計測を用いた実験計画を提案している[2]。ただし、実際の実験はまだ行われておらず、今後の研究の進展が期待されている。

本研究では、記憶の操作に代表される高次の認知機能に対応するとされる脳前部（前頭前野）[1][3]について、プログラム理解中の脳血流を NIRS（Near-infrared spectroscopy；近赤外分光法）により計測・分析する。

プログラム理解は前述の通り認知プロセスを含む知的活動であるため、前頭前野の活性がいくらか観測できることは確実と思われるが、一方、読解対象に含まれるデータ構造の複雑さや変数の量、使われ方といった特性の違いが定量的に観察できるかについては不明である。そこで、特に以下の仮説を検証する。

仮説：プログラムの読解難度や読解時に係る負荷の差を、読解中の脳血流計測によって定量的に観測することが出来る

プログラム理解行動の測定に関する従来研究では、思考の内容と行動を発話してもらってシンクアラウド法や、思考過程や感想をインタビューで調査する方法[4]などが用いら

* Takao Nakagawa, 奈良先端科学技術大学院大学

† Yasutaka Kamei, 九州大学大学院

‡ Hidetake Uwano, 奈良工業高等専門学校

§ Akito Monden, 奈良先端科学技術大学院大学

** Ken-ichi Matsumoto, 奈良先端科学技術大学院大学

れてきた。これらの手法では、発話自体が認知プロセスの一つであるため、あるいは実験終了後の記憶を頼りにしているため、純粹に理解のみに取り組む様子をリアルタイムに観察することは原理的に不可能である。

一方、脳血流計測においては装置取り付けの違和感を除いて、理解を阻害する要因はない。そのため、前述の仮説を確かめることによって、理解に取り組む様子のみを純粹に定量化することが可能となり、将来的にはプログラム理解に困難を感じている開発者の支援や、プログラム自体の理解容易性の定量的評価が可能となると考えられる。

2 実験

10名の被験者に難度の異なるプログラムを読んでもらい、その際の脳活動を NIRS によって記録する被験者実験を行った。被験者は奈良先端科学技術大学院大学の学生で、プログラミング経験年数3年以上、年齢は22～26歳であった。

2.1 実験方法

プログラム理解を行う方法には、コード中の単語や付属ドキュメントなどからトップダウンに理解する方法や、単純に実行過程を頭の中で追跡するボトムアップな方法（メンタルシミュレーション）などがある[5]。

これまでに、プログラムに対するメンタルシミュレーションが人間の短期記憶の操作と強く関係していることを示唆する報告が複数ある[6][7]。そこで本研究では、記憶の操作に代表される高次の脳活動をよく誘発できると考えられ、また、そのプロセスに比較的個人差が出にくいと思われるメンタルシミュレーションを調査の対象とした。

2.2 タスクと実験手順

実験タスクでは、C言語で17～32行の関数と、関数に渡す引数（前提条件）が書かれた問題用紙を読み、被験者がその動作をメンタルシミュレーションする過程を記録する。図1に、ソースコード（左）と前提条件（右）の例を示す。

関数中にはチェックポイントを表す目印が書かれた行が複数あり、読解がその行まで到達するたび、記録用紙にその時点の変数の値を記録する。チェックポイントはループ中にも含まれ、その場合は同じチェックポイントに複数回到達することになる。

実験者は記録用紙を随時チェックし、チェックポイントに到達するたび記録の正誤を

<pre>int func(int ***A, int N, int M, int L){ int i,j,k; int p; p=A[0][0][0];⁽¹⁾ for(i = 0; i < N; i++){ for(j = 0; j < M; j++){ for(k = 0; k < L; k++){ if(A[i][j][k] < p) p = A[i][j][k];⁽²⁾ } } } return p;⁽³⁾ }</pre>	<p>前提条件</p> <p>A[0][0][0] = 97, A[0][0][1] = 48 A[0][1][0] = 52, A[0][1][1] = 71 A[1][0][0] = 17, A[1][0][1] = 64 A[1][1][0] = 11, A[1][1][1] = 32 A[2][0][0] = 20, A[2][0][1] = 22 A[2][1][0] = 48, A[2][1][1] = 86</p> <p>N = 3, M = 2, L = 2</p>
--	--

図1 ソースコードと前提条件、解答の例

Quarifying Program Understanding Activities Based on Cerebral Blood Flow Measurement

確認し、被験者に伝える。被験者は記録が正しければそのまま読解を続け、誤っていたら前回のチェックポイントまで戻って読解を続ける。

タスクに用いる関数は理解が容易な3つ (easy タスク) と、easy タスクを難読化し、変数の数や処理構造を複雑にした3つ (hard タスク) の6つである。練習用タスクとして、同様の方法により p-easy タスクと、p-hard タスクの2つを用意した。本タスク用に作成した3つの関数と、練習タスク用に作成した関数の処理内容はそれぞれ、配列の最小値計算、配列内の数値の合算、配列内の特定要素の数え上げ (本タスク)、配列の最大値計算(練習タスク)である。

順序効果と学習効果による影響を防ぐため、本タスクの実施においては、easy タスクを先に提示される組と、hard タスクを先に提示される組の二種類を用意し (それぞれ5人)、また、本タスクにおいては処理内容が同じタスクを提示しないよう配慮した。

実験手順は以下の通りである。

- 手順 1. 実験内容の説明、装置の取り付け
- 手順 2. p-easy タスク、p-hard タスクの実施
- 手順 3. 休憩2分 + 本タスク1 (easy, もしくは hard) の実施
- 手順 4. 休憩2分 + 本タスク2 (本タスク1と異なる難易度) の実施
- 手順 5. 装置の取り外し・アンケートの実施

2.3 実験環境

実験には、日立製ウェアラブル光トポグラフィ WOT-220 を用いた。図2に、装置と、装着時の外観を示す。実験はすべて、被験者と実験者の2名のみが居る静かな部屋で、椅子に座った状態で行われた。被験者の体動を防ぐため、事前に下を向かないよう被験者に伝え、実験中にも、顕著な場合は口頭で頭を上げるように指示した。ソースコードと記録用紙は書見台に置き、できるだけ楽な姿勢で読めるよう、事前に椅子の位置や高さを調整してもらった。

2.4 分析方法

計測機器である NIRS は、脳血流の酸化傾向を測ることで脳活動を測定する装置である。すなわち、血中の酸化ヘモグロビンを表す oxy-Hb が多いほど、その部位での脳活動が活発であることを表す。

実験で得られるデータは 200ms ごとの oxy-Hb の値で、値が大きいほど脳の活動量が



図2 ウェアラブル光トポグラフィ WOT-220 (左) と装着の様子 (右)

多いことを示す。なお、計測値は計測開始時の oxy-Hb を $0[\text{mM}\cdot\text{mm}]^{\dagger\dagger}$ としたときの各時点での差分であるため、異なる被験者同士の比較や、異なる時期に計測された値を比較することは適切ではない。本稿では全被験者の平均値によるタスク間の比較は行わず、被験者ごとの実験結果を個別に比較する。

また、本稿では被験者ごとに計測される oxy-Hb の最大値と最小値の範囲を 0 から 1 に正規化（正規化 oxy-Hb ）し、分析する。

3 結果

3.1 実験結果の概略

全ての被験者について、easy タスク (296[s]) より hard タスク (1909[s]) のプログラム理解に時間を要した。また、手順 5. のアンケートでは、すべての被験者が hard タスクの読解のほうが難しく感じたという回答している。

3.2 仮説の検証

図 3 は、被験者ごとに難易度別の正規化 oxy-Hb の分布を表した箱ひげ図である。a から j が個々の被験者を表し、各々、左が easy タスク、右が hard タスクの分布を表している。縦軸は正規化 oxy-Hb を表し、値が高いほど脳が活動していることを表す。

図 3 からは、10 人中 8 人の正規化 oxy-Hb について、hard の中央値が easy の中央値よりも高かったことがわかる。このことは、仮説を支持し、プログラムに含まれる変数の数や更新頻度といったパラメータや、読解時にかかる負荷の差が、プログラム読解中の前頭前野の働きに影響したことを示している。すなわち、本結果は、読解中の脳血流計測により、プログラムの読解難度や読解時にかかる負荷の差を定量化できることを示唆している。

一方で、10 人中 2 人（被験者 e, g）が仮説と逆の結果を示した。過去の NIRS に関する研究では、10 人に 1 人程度の割合で逆傾向の被験者が含まれる可能性が指摘されており [8]、本実験の結果にもその影響があることが考えられる。例えば、被験者 e は事前の練習タスク中の脳血流でも、高難度問題で oxy-Hb が減少しており、

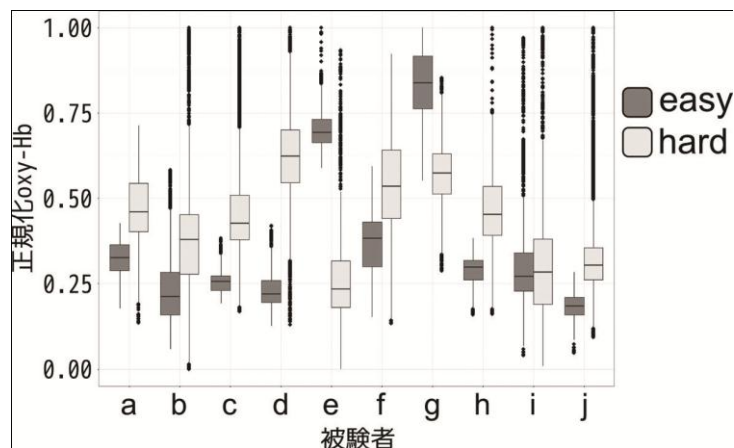


図 3 被験者ごとの脳血流量

^{††} M はリットルあたりの oxy-Hb 量を表すモル濃度(mol/L). mm は脳を通ったレーザの光路長.

Quantifying Program Understanding Activities Based on Cerebral Blood Flow Measurement

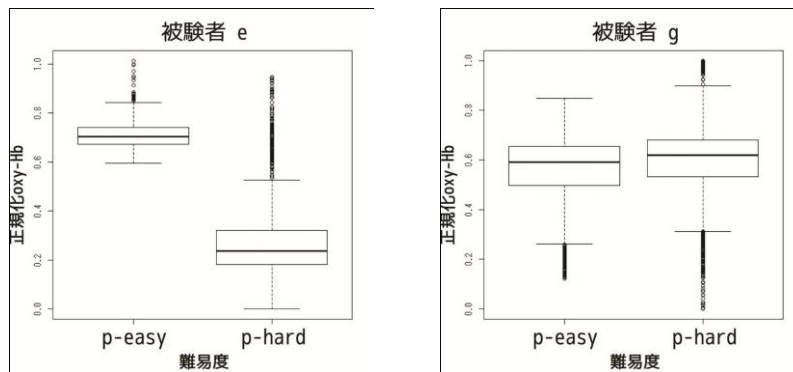


図4 被験者 e,g の練習タスクでの脳血流量の差異 (左:被験者 e, 右:被験者 g)
 一貫して逆傾向を示す被験者である可能性が高い (図 4 左). 一方で, 被験者 g は練習タスクにおいては高難度問題のほうが oxy-Hb の上昇が見られた (図 4 右).

また, 図 3 に着目すると, 正の傾向を示した被験者 8 名は (被験者 e, g と比較して) 難易度間で oxy-Hb の最小値の差が小さく, 最大値の差が大きかった. 一方で, 逆の傾向を示した被験者 2 名 (e, g) は難易度間で最小値に大きな違いがあった. この結果は, この被験者 2 名が他の 8 名と逆の傾向を示す被験者であることを示唆する.

また, タスク時における脳活動量の時系列変化を分析した. 被験者ごとにタスク時間を序盤, 中盤, 終盤に三等分した際の脳血流の変化を, 図 5 に示す.

図からは, hard タスクにおいて 10 人中 8 人で序盤より中盤のほうが正規化 oxy-Hb が高く, 10 人中 7 人では終盤より中盤のほうが正規化 oxy-Hb が高かったことが読み取れる. この原因として, 例えば中盤に最も被験者による記録の誤りが多かったことや, 終盤においては処理構造の把握が進んでいることなどが影響していると考えられる.

この結果は, プログラム理解中の負荷や対象の理解難度の脳血流計測による定量化が, 理解進度に応じた時系列分析へ応用できる可能性を示唆する.

4 おわりに

本論文では, プログラムを理解する際における, 理解対象のプログラムの難度や, 理解にかかる作業負荷を定量化するため, 記憶や想起に関する高次の認知プロセスに関連するとされる脳部位である前頭前野の脳血流を, プログラム理解中に計測する被験者実

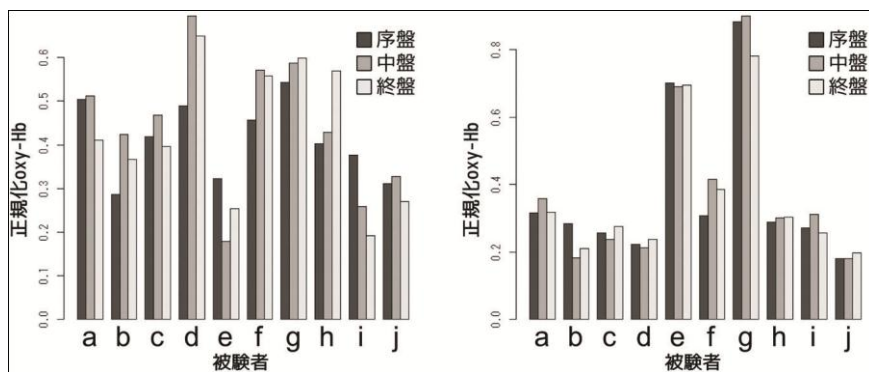


図 5 課題進捗度ごとの脳血流量の分布 (左:easy タスク, 右:hard タスク)

験を行った。実験では、変数の数・更新頻度が多く構造が複雑なタスクと、少ないタスクを用意し、その両方を読んだ際の脳血流量を連続的に取得した。

実験データからは、変数の数や更新頻度がより多く、構造が複雑なプログラムにおいて、前頭前野の脳血流量が上昇する傾向が確認できた。この結果は、プログラムの読解難度や読解時にかかる負荷の差を、読解中の脳血流計測によって定量的に観察することができる可能性を明らかにした。また、脳血流の時間変化を見ると、いくらかの被験者に共通の傾向が見られ、プログラム理解の進度に応じた難度の違いや作業負荷の差を時系列に分析することができる可能性が示唆された。

本研究の今後の発展として、他の認知プロセスと比較してプログラム理解がどのような位置づけになるか確認するため、プログラム理解の過程を自然言語文の読解や暗算などと比較する実験の実施が考えられる。また、一般に脳活動量の上昇が見られるとされる、読解や暗算などのタスクを併用することによって、逆傾向の被験者を事前に発見できる効果も期待できる。他に、何が直接脳活動に作用しているかを確認するため、より目的を絞った短いコード片を連続的に与える実験によって、影響を確認することも興味深いと思われる。

謝辞

本研究の一部は、文部科学省科学研究補助費（基盤 B: 課題番号 23300009）による助成を受けた。

参考文献

- [1] R. Cabeza and L. Nyberg: Imaging Cognition II: An Empirical Review of 275 PET and fMRI Studies, *J. of Cognitive Neuroscience*, Vol.12, Issue 1, pp.1-47, 2000.
- [2] R. Buckner and S. Petersen: What does neuroimaging tell us about the role of prefrontal cortex in memory retrieval?, *seminars in THE NEUROSCIENCES*, Vol 8, pp.47-55, 1996.
- [3] J. Siegmund, A. Brechmann, S. Apel, C. Kastner, J. Liebig, T. Leich and G. Saake: Toward Measuring Program Comprehension With Functional Magnetic Resonance Imaging, In *Proc. of International Symposium on the Foundations of Software Engineering (FSE '12)*, No. 24, 2012.
- [4] K. Ericsson and H. Simon: *Protocol analysis: Verbal reports as data*, MIT Press, Cambridge, 1984.
- [5] N. Pennington and B. Grabowski: The tasks of programming, In *Psychology of programming*, pp.45-62, London: Academic Press, 1990.
- [6] M. Nakamura, A. Monden, T. Itoh and K. Matsumoto: ueue-Based Cost Evaluation of Mental Simulation Process in Program Comprehension, In *Proc. of 9th IEEE International Software Metrics Symposium (METRICS2003)*, pp.351-360, 2003.
- [7] 石黒 誉久, 井垣 宏, 中村 匡秀, 門田 暁人, 松本 健一: 変数更新の回数と分散に基づくプログラムのメンタルシミュレーションコスト評価, *電子情報通信学会技術報告, ソフトウェアサイエンス研究会*, Vol.SS2004-32, pp.37-42, 2004.
- [8] 独立行政法人 国立特殊教育総合研究所: 脳科学と障害のある子どもの教育に関する研究, *課題別研究報告*, B-215, 2007.