



卒業研究報告書

令和4年度

研究題目

脳波の周波数成分の時系列変化を用いた
プログラム理解の判別

指導教員 上野秀剛 准教授

氏名 松山鼓

令和5年1月26日 提出

奈良工業高等専門学校 情報工学科

脳波の周波数成分の時系列変化を用いた

プログラム理解の判別

上野研究室 松山 鼓

ソフトウェア開発における主要な作業の1つであるプログラム理解を支援することで、効率的にプログラムを教えることが可能になる。そのためにはプログラムの理解状態を把握する必要があり、理解状態を把握する方法として脳波に着目した手法がある。プログラム理解に関する研究において、脳波は δ 波、 θ 波、 α 波、 β 波、 γ 波の5つに分類されることが多い。その中でも α 波、 β 波にプログラム理解時の特徴が強く現れるとされている。周波数帯は δ 波、 θ 波、*Low* α 波、*High* α 波、*Low* β 波、*High* β 波、*Low* γ 波、*Mid* γ 波の8種類に細分化でき、細かな分類の中にもプログラム理解に関する特徴があるとされている。また、脳波を用いた分析においては時系列分析が有効とされており、 α 波、 β 波に対して時系列分析を行った研究も存在する。しかし、複数の周波数成分に対して時系列分析は行われていない。そこで本研究では、プログラム理解を行う開発者の脳波に対して複数の周波数成分の時系列分析を行い、機械学習を用いることで理解の有無を自動で判別できるか評価する。評価対象として、先行研究で取得されたプログラム理解タスク中の脳波データを用いる。分析手順として、初めに8種類の周波数成分に対して時系列への分割を行う。1分割(分割なし)から10分割まで、分割数 d ($d=1, 2, \dots, 10$) を変更した複数のデータセットを作成し、FFTを用いて分割データごとに周波数成分のパワースペクトルを求める。データセットにはプログラム理解タスクの正否をラベルとして付与する。次に、作成したデータセットをRandom forestによって機械学習させ、新規のデータに対してプログラム理解の有無を予測させる。最後に、周波数成分を5種に分類した際のデータセットも同様に作成し、機械学習、複数の周波数成分時との比較を行い、評価を行う。実験の結果、分割なしの状態においては約68.9%程の精度だったのに対し、分割数を上げると3分割以上では精度が向上し、最大86.5%の精度が確認できた。また、周波数成分が8種類の場合においては、最大83.8%、平均で77.9%の精度で、周波数成分が5種類の場合においては、最大で86.5%、平均で77.0%の精度で予測ができた。以上の結果は、脳波を複数の時間帯に分割して得られる周波数成分からプログラム理解状態を高精度で予測することが可能であることを示している。また、時系列ごとの分析により、周波数成分5種、8種の双方においてタスクを一定の時間毎に分割した場合、10分割時に1番目、5番目から8番目の各周波数成分が精度に強い影響を及ぼしていた。このことから、プログラムを理解できた被験者は1)タスク開始と同時に問題を見て大まかな全体像を理解し、2)プログラムを読み進めて段階的に理解を進めていき、3)プログラムを読み終わり細部も含めて理解し、4)タスクへの回答を確認する、という工程を経ていると考えられる。

目次

1	はじめに	1
2	関連研究	3
3	脳波	5
3.1	測定方法	5
3.2	周波数帯域	5
4	実験	7
4.1	実験環境 [8]	7
4.2	タスク	7
4.3	実験の手順 [8]	9
4.4	分析	9
5	結果と考察	11
5.1	パラメータ設定による予測精度の変化	11
5.2	時間軸ごとの重要度	13
5.3	プログラム理解中の周波数成分	14
6	おわりに	17
	謝辞	18
	参考文献	19

1 はじめに

プログラム理解はソフトウェア開発における主要な作業の1つである。例えば、プログラム教育現場などにおいて、学習者がプログラムを理解しているのかどうかを測定することで、教育者は学習者がプログラムを十分に理解できているかを確認でき、学習者が十分に理解するまで教えることが可能になる。これらは口頭でも行える作業だが、学習者の中には理解していないことを報告しにくい人もいる可能性があり、プログラム理解に対する支援を充分に行えない状況が存在する。そこで、プログラム理解状態を自動的に判別するシステムがあれば、指導者は学習者への効率的なプログラム理解支援が可能になる。しかし、プログラミング作業には思考作業が多く含まれる為、指導者側が外観から理解状態を把握することは難しい。

プログラム理解状態を把握するための方法として、脳活動や視線情報などの生体情報に着目する手法がある。生体情報の1つである脳活動は、思考状態や心理状態と関係があることが知られており、プログラム理解を測定するための様々な研究で用いられている [1, 2]。脳活動を計測する為の手法の中でも、脳波はミリ秒単位で記録が行え、時系列ごとの細かな変化を把握することが可能なことから、時間分解能が高く、プログラム作業中の理解状態を判別することに適している手法だと考えられる。脳波は国際脳波学会によって定められた周波数成分ごとに名称が付けられており、被験者の心理状態によって各周波数成分の強さが変化することがわかっている。リラックス時などの安静状態では α 波が、ストレスがある状態では β 波が出現することが知られている [3]。このような被験者の集中や思考状態によって脳波が変化するという特徴を用いることで、プログラムの理解状態を識別できる。また、脳波はタスクを理解しているのか、理解していないのかなど、タスクに対する思考状態によって逐次変化するため、タスクを一定の時間軸ごとに区切るなどを行い、その時間軸ごとに時系列分析することで、理解状態の把握に必要な情報を得ることができると考えられる。

プログラム理解時の脳波を計測した従来の研究では、脳波の複数の周波数成分に着目し、Low β 波、High β 波、Low γ 波、Mid γ 波がプログラム理解の有無を把握する指標となる可能性があることがわかっている [4]。また、プログラムのプログラム理解の有無によって α 波と β 波の時系列変化に有意な差があることがわかっている [5]。一方で、 α 波、 β 波以外の複数の周波数成分において時系列変化に有意差があるのかは明らかにされておらず、時系列変化の情報によって α 波、 β 波以外に理解状態の把握に対してより重要な情報が含まれている可能性がある。従って、機械学習を用いることによって、 α 波、 β 波以外に理解状態の把握により重要な周波数成分があるかどうか確かめる。

本研究はプログラム理解を行う開発者の脳波に対して複数の周波数成分の時

系列分析を行い、機械学習を用いることで理解の有無を自動で判別できるか評価する。機械学習では、タスク中の周波数成分を等間隔に複数の領域に分割し、FFTを用いて正規化したデータを入力とする。また、機械学習を用いることで理解の有無の判別にどの周波数成分が重要か分析する。分析によって、予測に必要なデータはどこの時系列のデータなのか、どの周波数成分を用いて予測しているのかを知ることができ、プログラム理解の過程を考察することが可能になる。

以下、2章では関連研究について説明し、3章では脳波についての説明する。4章では実験内容について説明し、5章では実験結果の考察、6章で本研究での結論を述べる。

2 関連研究

プログラムの理解状態の定量的な評価に生体情報を利用した研究が数多く存在する。生体情報とは視線、脳波、心音などの生体活動を調節する体内活動に関する情報のことであり、その中でも近年、脳波計測装置の普及に伴い、プログラム理解と脳計測を組み合わせた研究が数多くある。

Siegmundらは脳の血流変化を測定する装置であるfMRIを用いてプログラム理解時の脳活動を観測した[1]。脳は脳領域ごとに役割が分かれており、注意や言語理解をしている際など、思考によって特定の脳領域が活性化する。fMRIを用いることでプログラムを理解している際の血流変化を測定し、どの脳領域が活性化するのかを測定することができる。Siegmundらの研究の結果、プログラムを理解する時に、被験者の問題解決、記憶、および文章理解に関する脳領域が活性化することが示されている。fMRIを用いた計測はどの脳領域が活性化しているのかを明確に明らかにできる為、プログラム理解と特定の脳部位の関係性について正確に明らかにすることができる一方、被験者が装置の中に横たわって測定する必要があり、被験者の体動や思考を強く制限する。その為、教育現場や仕事場などの実際のプログラム環境において理解状態を計測し、プログラム理解を支援することは難しい。そこで、fMRIでの計測と比べて体動の制限が少ない脳波計測装置を用いた研究が注目されている。

Martinらは脳波計測装置を用いて、プログラム理解における脳活動を記録した[6]。脳波を計測することで、異なる種類の短いソースコードに対する脳の反応が異なるかどうかを調査した。Martinらの研究の結果、認知的負荷が高い場合には、 α 波と θ 波のパワースペクトルが増加することが示されている。また、他の研究において α 波と θ 波には負の相関があることが確認されている。Klimeschは、 α 波が記憶能力に応じて変化することを観測している[7]。 α 波の中でもHigh α 波が意味記憶に対して、Low α 波が集中している時に反映されることが示されている。石田はバグを探しているプログラマの状態を知るために視線と脳波の分析を試みた[8]。プログラマの状態を把握するために視線や脳波の計測を行う研究は多く行われていたが、視線と脳波を組み合わせた研究は行われていなかった。プログラマの状態をプログラムの動作を理解しているか、バグを正しく判断出来るかの2つに分け、2つの状態時の分析を行った。分析の結果、プログラム動作を理解している時には α 波が増幅し、バグを判断出来た時には β 波が増幅することが確認できた。また石田らは別研究[5]において、視線と脳波に対して同期した時系列分析する為に、デバッグを行うプログラマの理解過程を時系列に計測することを目的に、脳波特徴の時系列分析を試みた。実験の結果、プログラムの動作を理解をしている被験者はタスク開始時と終了時の α 波と β 波が他の時系列の脳波と比較して有意に高く、バグを正しく判断している被験者はタスク開始時

と終了時の α 波が他の時系列の脳波と比較して有意に高くなることが示されている。このように、脳波は被験者の状態によって逐次変化する為、リアルタイムでの支援を行う為には時系列分析が有効とされている。

また、 α 波、 β 波、 θ 波以外の周波数成分とプログラム理解の関係を調査する研究がある。また、柴田は脳波の各周波数成分において詳細な心理状態との関連を示すために、 α 波、 β 波を Low, High, γ 波を Low, Mid に分割し、複数の周波数成分を用いてプログラム理解状態の脳波の分析を行った [4]。被験者を正答者、未回答者、誤答者に分類し、分析を行った結果、Low β 波、High β 波、Low γ 波、Mid γ 波において3つの状態に有意差があることが示されている。一方で、柴田の研究では複数の周波数成分に対して分析を行っているものの、タスク全体の脳波に対してのみ分析しており時系列分析が行われていない。

脳波などの生体情報を用いて機械学習を行った研究も存在する。Fritzらは視線、皮膚電気活動、脳波を計測し、被験者がタスクに対して困難と感じているかどうかを機械学習を用いて推定した [9]。結果として、タスクの難易度(易しい/難しい)を 64.99 % の精度、64.58 % の再現率、被験者が初めて行うタスクの場合 84.38 % の精度、69.79 % の再現率で予測できた。脳波などの生体情報は、プログラム理解中の脳波に特徴があるため、プログラムの理解状態を推測する機械学習に有用である。

本研究ではプログラム理解中の複数の周波数成分に対して時系列分析を行い、機械学習を用いることでプログラム理解状態をどれ程の精度で予測できるのか、また、予測にはどの周波数成分が重要なのか検証する。

3 脳波

3.1 測定方法

脳波とは，脳から生じる電気活動を縦軸，時間を横軸にとって記録したものである [10]. 脳波は頭皮上に装着した電極から計測される．電極の配置は図1に示す国際式10-20電極法に則って行う [11]. 国際式10-20電極法では頭皮が10%もしくは20%の等間隔で区切られており，検査や研究の用途によって電極配置位置を決定する．

脳波には基準電極導出法と双極導出法の2種類の導出法がある．基準電極導出法では，脳電位の電場内に装着した探査電極と，電場外に装着した基準電極の2つの電極の電位差として脳電位を測定する．探査電極は図1に従って測定したい部位に装着する．双極導出法では，基準電極を用いず，2つの電極を脳電位の電場内において記録を行う．2つの電極のそれぞれがとらえた脳電位の差が記録される．優劣な背景成分を除去して部位差を強調する目的で使うことが多い [12].

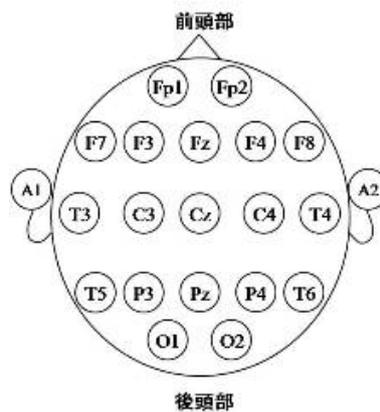


図1 国際式10-20電極法

3.2 周波数帯域

脳波は異なる周波数をもつ波形を合成したものであり，周波数帯域ごとに異なる特徴をもつ成分として分類される．周波数成分を抽出するためにFFT (Fast Fourier Transform) やSTFT (Short Time Fourier Transform) が多くの研究で用いられる．脳波の周波数成分の代表的な分類は以下の通りである．

- δ 波：0.5~4Hz
- θ 波：4~8Hz
- α 波：8~14Hz
- β 波：14~30Hz

表1 脳波の分類

周波数成分	周波数帯域	心理状態
δ波	0.5~2.75[Hz]	ノンレム睡眠, 無意識
θ波	3.5~6.75[Hz]	直観的, 創造的, 想起, 空想, 幻想, 夢
Low α波	7.5~9.25[Hz]	リラックス, 平穏, 意識的
High α波	10~11.75[Hz]	リラックスしているが集中している, 統合的
Low β波	13~16.75[Hz]	思考, 自己及び環境の認識
High β波	18~29.75[Hz]	警戒, 動揺
Low γ波	31~39.75[Hz]	記憶, 高次精神活動
Mid γ波	41~49.75[Hz]	視覚情報処理

- γ波: 30Hz以上

δ波やθ波は睡眠状態にあるときに出現する。α波はリラックス時や、ストレスが少ない時などの安静状態に出現する。他の周波数帯域の波と比較して振幅も連続性も最も高い。また、α波は開眼すると大幅に減少し、閉眼すると再び出現する。この現象を一般的にα波減衰と呼ぶ。β波は緊張や不快な感情を抱いている時など、ストレスが強い時に出現する。γ波は、不安で興奮している時に出現しやすい[12, 3].

周波数帯域は更に細かい分類が定義されており、その特徴を表1に示す[13]. このように各周波数帯域が被験者の心理状態によって変化する特徴を持つ。これまでの研究においてα波、β波はプログラム理解に関係があることが明らかになっている[8]. また、周波数帯域を細かく区分した際にはLow β波、High β波、Low γ波、Mid γ波がプログラム理解と関係があることが示されている[4]. 時系列分析を行った際においても、これらの周波数帯域がプログラム理解時の特徴として変動することが期待できる。

4 実験

本研究では先行研究[8]が行った実験で取得された脳波データを用いて分析する。先行研究の実験環境を4.1節に、タスクを4.2節に、実験手順を4.3節に示す。

実験は日本語とJavaで記述されたプログラム課題を被験者に提示し、処理内容を理解する間の脳波を計測する。被験者は奈良工業高等専門学校の学生5人で、年齢は19歳から20歳、全員がJavaによるプログラミングの基礎講義を受講済みである。

4.1 実験環境 [8]

実験は被験者1名と実験者2名のみが居る静かな部屋で実施する。体動によるアーチファクトを抑えるために、ひじ掛け・足置きを備えた椅子に座り、頭部と体をできるだけ動かさないように指示する。

実験に使用するのは、脳波計測装置、タスク提示用PC、脳波計測用PC、記録用PCである。脳波計測装置にはナノテックイメージ社製Nexus-10 MARK IIを用いる。本計測装置は計測周期は256Hzであり、計測された脳波はBluetooth経由で脳波計測用PCに転送され、CSVファイル形式で出力される。計測する際には本計測装置に対応したソフトウェアであるBioTrace+ Softwareを用いる。実験装置を図2に示す。



図2 実験装置

4.2 タスク

提示するプログラム課題の処理内容を理解するタスクを被験者1人につき16問与える。実験用に作成したタスク提示ツールを用いて、被験者に日本語で記述されたプログラムの仕様と対応するJavaのソースコード1組を提示する。各タスクは動作理解ステップとバグ判断ステップで構成される。動作理解ステップでは、被験者にプログラムの仕様とソースコードを提示し、動作を理解しているかどうかの質問を行う。質問には口頭で回答してもらい、事前に用意した回答と一致

表2 タスク一覧

	難易度	仕様
1	easy	階乗の計算
2		最大値検索
3		表数判定
4		中央値検索
5		累乗検索
6		2つの数値の入れ替え
7		指定した文字列が含まれているか判定
8		文字列を反転させる
9	difficult	ハノイの塔
10		経路数を求める
11		順列を全列挙する
12		組合わせを漸化式から求める
13		支払う硬貨の組み合わせを求める
14		文字列の組み合わせを求める
15		雲の軌道予測
16		最小公倍数と最大公約数

していれば動作を理解している、一致していなければ理解していないとみなす。バグ判断ステップでは、被験者に動作理解タスクで見たソースコードの一部を改編したソースコードを見せる。被験者には改編後のソースコードが仕様を満たすかどうかを回答してもらう。回答の正誤は被験者には伝えず、制限時間は1分とする。

本実験では、プログラムに対する理解の有無が同程度の件数計測できるように、タスクの難易度と制限時間を調整する。タスクの難易度は低難易度と高難易度をそれぞれ8件用意する。低難易度はmainメソッドのみからなり、1重の繰り返し文や条件分岐で構成された理解が容易と思われるソースコードを使用する。高難易度のソースコードは複数メソッドの使用や再帰構造を持ち、制限時間以内での理解が難しいと期待できる複雑なアルゴリズムを使用する。制限時間はそれぞれのステップごとに設定し、低難易度の理解には十分で、かつ高難易度の理解に不十分であることを予備実験によって確認している。また、各タスクを提示する順番は、順序効果を考慮し、カウンターバランスを行う。表2にタスク一覧を示す。

なお、先行研究[8]の2つのステップの内、本研究では動作理解ステップから得たデータのみを分析に使用する。

4.3 実験の手順 [8]

実験の順序を以下に示す.

1. 実験の説明

実験概要, 脳波・視線計測時の注意点を説明する.

2. 装置の装着・設定

脳波計測装置 NeXus-10 MARKII を被験者に, 視線計測装置 Eye Tracker 4C を PC に装着し, 設定する.

3. 練習問題

1タスクの流れを被験者に確認してもらうために, 練習タスクを実施してもらう.

4. 動作理解タスクの実施

ディスプレイにタスクを2分30秒間提示し, プログラムの動作を理解してもらう.

5. 解答

被験者には動作を理解した時にコントローラのボタンを押して答えを口述してもらう.

6. 変更後タスクの実施

ディスプレイに変更後のタスクを1分間提示し, 仕様を満たすか判断してもらう.

7. 解答

被験者には仕様を満たすか判断できた時にコントローラのボタンを押して答えを口述してもらう.

8. 全タスクの実施

1タスク当たり手順4~7を1回を行う. 全タスクである16タスクを行うため, 手順4~6を16回繰り返す.

4.4 分析

本研究では複数の周波数成分に対して時系列分析する為, 各被験者の1タスクごとに, タスク全体の脳波データを時系列ごとに分割する. 分割の際, 分割数によって予測精度に変化があるのか, どの時系列の脳波データがプログラム理解の予測に重要になっているのかを分析する為に, 1分割(分割なし)から10分割まで, 分割数 d ($d=1, 2, \dots, 10$) を変更した複数のデータセットを作成する.

分割数の異なる複数のデータセットそれぞれに含まれる i 番目の脳波 $BW_i (i=1, 2, \dots, d)$ に対してFFTを用いて、表1に示した周波数成分ごとのパワースペクトル8種類 ($\delta_i, \theta_i, L\alpha_i, H\alpha_i, L\beta_i, H\beta_i, L\gamma_i, M\gamma_i$) を求める。脳波の個人差を考慮し、抽出されたパワースペクトルに対して、タスク全体のパワースペクトルの合計を求めることで、被験者ごとのパワースペクトルの平均値を導出する。各周波数成分ごとのパワースペクトルを被験者の全てのパワースペクトルの平均値によって割ることで正規化を行い、得られたデータをRandom forestによって機械学習させる。

正解集合は実験時に集計した、被験者の自己申告による理解の有無とする。Leave one out法を用いてすべてのタスクデータに対して理解状態の有無をYesかNoで予測し、データセットごとの精度を算出する。

Random forestのパラメータ設定は決定木モデル数 n_{tree} は500, 使用メトリクス数 m_{try} はRandom forestのモデルに対してチューニングを行った際に、エラー率の少なかつた1から16までの値を用いて分析する。また、周波数成分の細かな分類がプログラム理解状態の推定精度に影響を与えるか確かめるため、 α 波、 β 波をLow, Highに、 γ 波をLow, midにそれぞれ分類しなかつた際の5種類 ($\delta_i, \theta_i, \alpha_i, \beta_i, \gamma_i$), 分割数 $d (d=1, 2, \dots, 10)$ の周波数成分を用いたデータセットを作成し精度を比較する。Random forestにはR言語のrandomForestパッケージを、その他の分析にはdplyr, pforeachパッケージを用いた。

図3に分析の流れをまとめたフロー図を示す。

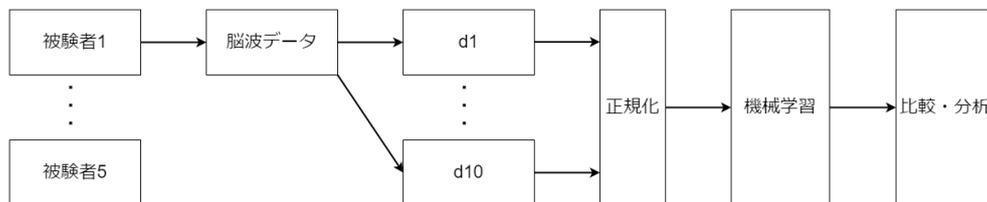


図3 分析の流れ

5 結果と考察

実験の結果、被験者1人につき16タスク、5人×16タスクの合計80件のデータを取得した。その中から異常値を含む6件を削除した74件の脳波データを用いて分析する。データセットは脳波を8種類の周波数成分とした $d=1\sim 10$ の10個と、脳波を5種類の周波数成分とした $d=1\sim 10$ の10個の計20個について評価する。

5.1 パラメータ設定による予測精度の変化

周波数成分を細かく分類した際のデータセットに対する予測精度を図4に示す。図の横軸は $mtry$ 、縦軸は予測精度、線の種類は各分割数を表す。 $d=1\sim 5$ を図(a)に、 $d=6\sim 10$ を図(b)に示している。その結果、周波数成分を細かく分類したデータを入力としたとき、 $div1$, $div2$ が精度が最も低く、 $div8$, $div9$ の時に最大精度が出ているものの、 $div3$ 以降の精度には大きな変化がなかった。補足情報として、最大精度は $div8$ の $mtry1, 4, 5$, $div9$ の $mtry7, 16$ で約83.8%、 $div1$ の $mtry1, 2, 3, 4, 9, 10$ の時に最小精度68.9%、全体の平均として約77.9%の確率でプログラムの理解状態を予測出来るという精度が出た。本研究では、タスクを理解をしていた被験者と理解をしていなかった被験者の比率は、およそ6:4となっている。したがって、本研究の予測結果は周波数成分に対して時系列分析を行ったデータが、プログラム理解の予測に対して重要な情報であったと言える。

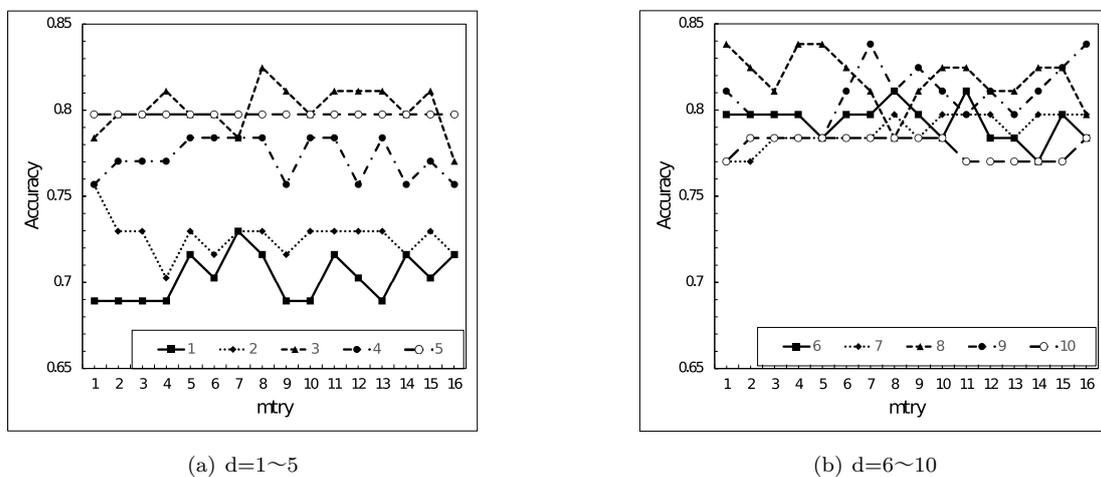
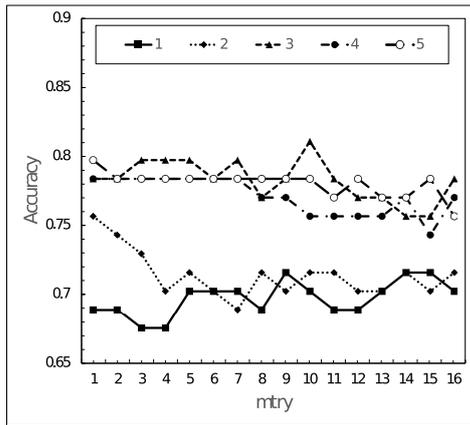
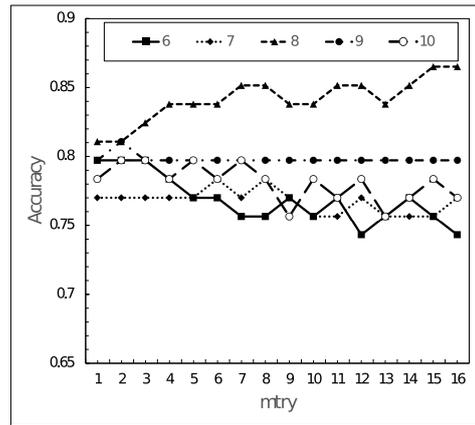


図4 8種類の周波数成分を用いた際の予測精度

次に、周波数成分を細かく分類しなかった際のデータセットに対する予測精度結果を図5に示す。グラフの要素は図4と同様のものである。その結果、周波数成分を細かく分類しなかったデータを入力としたとき、4の時と同様に $div1$, $div2$ が精度が最も低く、 $div8$, $div9$ の時に最大精度が出ている。また、 $div8$ は $mtry$ の値が大きくなるにつれて精度も上昇している。補足情報として、最大精度は $div8$ の $mtry15$,



(a) d=1~5



(b) d=6~10

図5 5種類の周波数成分を用いた際の予測精度

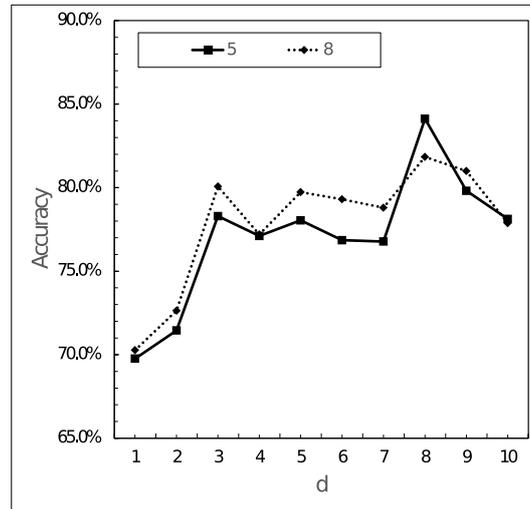


図6 分割数dごとの精度平均

16で約86.5%、div1のmtry3, 4の時に最小精度約69.0%、全体の平均で約77.0%という精度が出た。特に、div8でmtry15, 16の時の最大精度である約86.5%は本研究において最も高い精度となっている。しかし、予測精度を周波数成分を細かく分類して学習した際と比較すると、平均値の変化が少ないことから、周波数成分を細かく分類するかしないかに大きな変化はないと考えられる。

次に、脳波データの分割数dが精度に与える影響を分析する。図6に周波数成分8種と5種それぞれにおける、分割数dごとの予測精度の平均を示す。図の横軸は分割数d、縦軸は予測精度の平均値、線の種類は周波数成分の種類数を表す。

分割数1(分割なし)、分割数2の際には精度が極端に落ちていることが確認できる。分割数が3以上の場合における精度は、周波数成分8種を用いた際には分割数3以降の予測精度は分割数1, 2の精度よりも高く、大きな変化がないことも

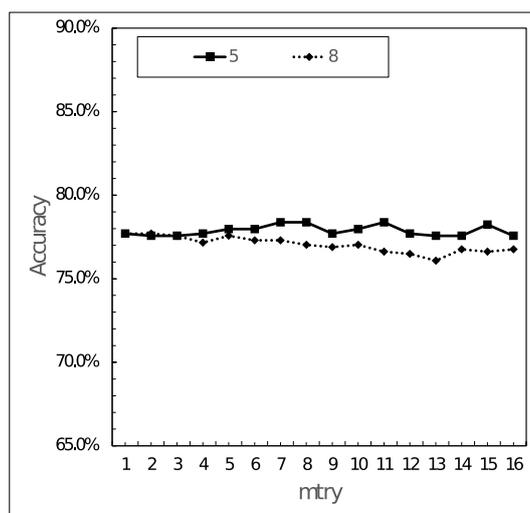


図7 mtry ごとの精度平均

確認できる．一方で，周波数成分5種の分割数8は他と比べても予測精度が高い．

分割数が1, 2の時に精度が低くなった理由として，分割数が3未満の時に算出される周波数成分のパワースペクトル（たとえば α_1 ）データには予測に必要なデータが多く含まれていることが要因だと考えられる．同様に，周波数成分5種の分割数8の予測精度が高い理由として，8分割した際に算出される周波数成分のパワースペクトルがプログラム理解の有無と強い関係を持つことが考えられる．

次にmtryの変更による予測への影響について考える．表7にmtryごとの予測精度の平均を示す．図の横軸はmtry，縦軸は予測精度の平均値，線の種類は周波数成分の種類数を表す．

周波数成分8種と5種のいずれにおいてもmtryごとの予測精度の平均値に差異が少ないことから，mtryは予測精度への影響は少ないと考えられる．mtryは予測に用いるメトリクス数を表すことから，多くの周波数成分を用いて予測を行っているのではなく，一部の周波数成分が予測結果に強い影響を与えていることが考えられる．

以上の結果から，プログラムを理解した際に表れる特徴は全体の中の一部の時間軸，特に8分割数した際に表れる時間軸のデータ，そしてそのデータ中の一部の周波数成分に大きく表れていると考えられる．

5.2 時間軸ごとの重要度

時間軸を複数に分割した各メトリクスについて，それぞれが予測精度に与える影響を確認する．メトリクス数の変化が予測に大きな変化を与えないこと，周波数成分を5種に分類した際の予測精度として8分割の際に最も高い精度が表れ

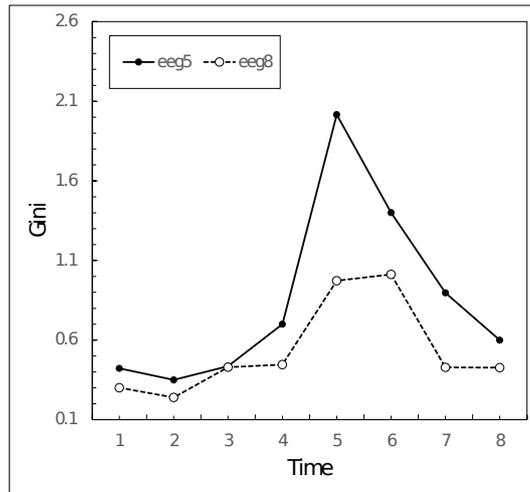


図8 時間軸による重要度の変化(分割数d=8)

たことから、分割数8, メトリクス数16のパラメータ設定を用いて予測した際の結果から考察を行う。Random forestによって作成したモデルからは、予測に使用したメトリクスが予測精度に与える影響を示すGini係数が得られる。同じ時間帯の周波数成分ごとのGini係数の平均値を図8に示す。横軸が時間、縦軸がGini係数を表す。図8から、周波数成分5種, 8種の双方において8分割した時間帯の5番目, 6番目の各周波数成分が最も精度に影響していた。タスク開始からタスク終了までの間において、半分を少し過ぎた頃の周波数成分データがプログラム理解の予測に対して最も重要な情報だと考えられる。

時間帯と予測精度の関係を確認するために、メトリクス数は変更せず、分割数のみ10に変更したデータから考察を行う。時間軸ごとの重要度の変化を図9に示す。周波数成分5種, 8種の双方において1番目, 5番目から8番目の各周波数成分が精度に強い影響を及ぼしている。このことは、プログラムを理解できている被験者の特徴がプログラム理解タスクの開始直後, および中盤以降に現れることを意味する。プログラムを理解できた被験者は1)タスク開始と同時に問題を見て大まかな全体像を理解し, 2)プログラムを読み進めて段階的に理解を進めていき, 3)プログラムを読み終わり細部も含めて理解し, 4)タスクへの回答を確認する, という工程を経ており, 1),3)の工程で特徴が現れたと考えられる。

5.3 プログラム理解中の周波数成分

プログラム理解状態の予測に対して、最も重要度が高い周波数成分について考察を行う。図10に、周波数成分を α 波から γ 波までの5種について、分割数8と分割数10のGini係数を示す。図10(a)から周波数成分5種の場合において分割数が8と10の場合どちらにおいても α 波, β 波のGini係数が高い。また、図10(b)の周波

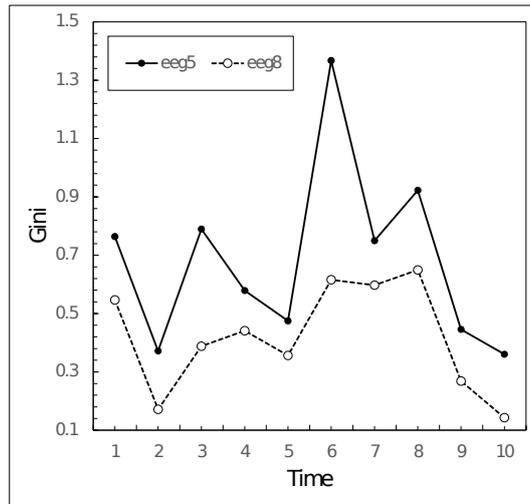
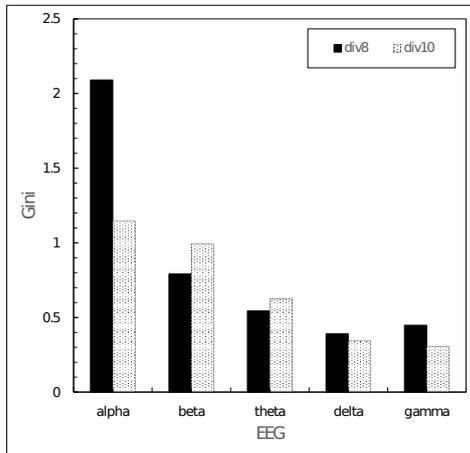


図9 時間軸による重要度の変化(分割数d=10)

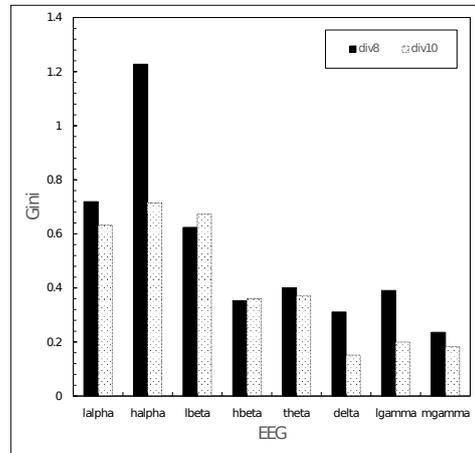
数成分8種の場合においても同様のことが確認でき、その中でも High α 波の Gini 係数が特に高くなっている。従って、どの時間軸においても α 波、 β 波はプログラム理解の予測に重要な情報であるということが言える。

しかし、上記結果は平均なので、特定の時間軸において他の周波数成分が予測に対して重要な情報になっている可能性がある。そこで、プログラム理解の予測に対して最も重要な時間軸になっていた、8分割時において前から5番目、10分割時には前から6番目のデータについて、周波数成分の Gini 係数がどのようになっているのか確認を行う。表3に8分割時の5番目、10分割時の6番目のデータの各周波数成分ごとの Gini 係数を示す。表の(5)は5番目の時間帯、(6)は6番目の時間帯を表す。分割数が8の場合について、周波数成分5種で α 波が最も高く、次点で β 波が高い。周波数成分8種で High α 波が最も高く、5種の時と同様に α 波、 β 波の Gini 係数が高くなっている。また、分割数が10の場合について、周波数成分5種で β 波、次点で α 波の Gini 係数が高い。周波数成分8種では、Low α 波、Low β 波において Gini 係数が高い。以上の分析から、 α 波と β 波がプログラム理解の予測において重要な指標であり、 α 波、 β 波以外の周波数成分は予測精度の向上に重要ではないことが明らかになった。

本結果は、時間帯を分割した分析を行っていない先行研究[4]の結果とは異なり、時系列ごとに分析したデータでは γ 波に特徴が表れないという結果になった。今後、周波数成分とプログラム理解の予測に関する分析を進めるために、時系列ごとの周波数成分のパワースペクトルが示す意味について、より深く分析する必要がある。



(a) 周波数成分5種



(b) 周波数成分8種

図10 周波数成分ごとの Gini 係数

表3 特定の時間軸における Gini 係数

d=8		d=10	
周波数成分	Gini 係数	周波数成分	Gini 係数
α (5)	6.848	α (6)	1.249
β (5)	1.891	β (6)	4.398
θ (5)	0.292	θ (6)	0.585
δ (5)	0.297	δ (6)	0.298
γ (5)	0.754	γ (6)	0.341
<i>Low</i> α (5)	1.783	<i>Low</i> α (6)	1.341
<i>High</i> α (5)	2.854	<i>High</i> α (6)	0.671
<i>Low</i> β (5)	1.118	<i>Low</i> β (6)	1.139
<i>High</i> β (5)	0.955	<i>High</i> β (6)	0.778
θ (5)	0.242	θ (6)	0.447
δ (5)	0.287	δ (6)	0.219
<i>Low</i> γ (5)	0.339	<i>Low</i> γ (6)	0.250
<i>Mid</i> γ (5)	0.207	<i>Mid</i> γ (6)	0.079

6 おわりに

本研究では、プログラムのプログラム理解状態を複数の時間帯に分割した脳波から得られる周波数成分を特徴量として機械学習で予測させることを目的として、プログラム理解時の複数の周波数成分に対して分析を行った。実験データについては、過去研究 [8] のデータを用いた。被験者のタスクごとの脳波データを、1分割(分割なし)から10分割まで、一定の時間ごとに分割し、周波数成分ごとのパワースペクトルを算出した。周波数成分が5種類($\delta, \theta, \alpha, \beta, \gamma$)と8種類($\delta, \theta, Low\alpha, High\alpha, Low\beta, High\beta, Low\gamma, Middle\gamma$)とした場合のデータセットを作成し、それぞれを用いた場合の予測精度を比較した。予測にはRandom forestを用いた。

分析の結果、周波数成分が8種類の場合において、時間帯の分割数が8,9の際の83.8%が最も予測精度が高く、分割数が1の場合の68.9%が最も低く、平均で77.9%の確率でプログラムの理解状態を正しく予測できた。また、周波数成分が5種類の場合において、時間帯の分割数が8の際の86.5%が最も予測精度が高く、分割数が1の場合の68.9%が最も低く、平均で77.0%という精度だった。以上の結果は、脳波を複数の時間帯に分割して得られる周波数成分がプログラム理解状態を高精度で予測できることを示している。また、周波数成分が8種類と5種類の場合で予測精度に違いは見られないことから、周波数成分を細かく分類する必要はないと考えられる。また、時系列ごとの分析により、周波数成分5種、8種の双方において10分割時に1番目、5番目から8番目の各周波数成分が精度に強い影響を及ぼしていることがわかった。このことから、プログラムを理解できた被験者は1)タスク開始と同時に問題を見て大まかな全体像を理解し、2)プログラムを読み進めて段階的に理解を進めていき、3)プログラムを読み終わり細部も含めて理解し、4)タスクへの回答を確認する、という工程を経ており、1),3)の工程で特徴が現れたと考えられる。更に、周波数成分の分析により、 α 波と β 波がプログラム理解の予測において重要な指標であるということがわかり、 α 波、 β 波以外の周波数成分は予測精度の向上に重要ではないことが明らかになった。

本研究の分析結果から、プログラム理解を行う開発者の脳波に対して複数の周波数成分の時系列分析を行い、機械学習を用いることで理解の有無を自動で判別できることが判明した。しかし、各時系列ごとの周波数成分のパワースペクトルが示す意味について、詳細に分析することが出来ていない。また、プログラムを理解できた被験者の理解過程については予測することが出来たが、理解できていない被験者の理解過程は予測することが出来なかった。今後は、理解した被験者と理解していない被験者の比較や、時系列ごとの周波数成分について、全ての時系列で比較を行う必要があると考えられる。

謝辞

本論文の執筆，研究を進めるにあたって，多くの方々にお力添えいただきました。指導教員である上野准教授には，お忙しい中，研究に対する助言，論文の添削など多くの助力をいただきました。査読教員である山口賢一教授には，研究に対する大変貴重な意見をいただきました。深く御礼申し上げます。本当に有難うございました。

参考文献

- [1] J. Siegmund, C. Kanster, S. Apel, C. Parnin, A. Bethmann, T. Leich, G. Saake, A. Brechmann: "Understanding Understanding Source Code with Functional Magnetic Resonance Imaging", Proceedings of the International Conference on Software Engineering (ICSE), pp.378-389 (2014).
- [2] 中川尊雄, 亀井靖高, 上野秀剛, 門田暁人, 鵜林尚靖, 松本健一: "脳活動に基づくプログラム理解の困難さ測定", コンピュータソフトウェア, Vol.33, No.2, pp.78-89,(2016).
- [3] 宮田洋, 藤澤清, 柿木昇治, 山崎勝男:"新生理心理学-生理心理学の基礎":北大路書房(1998).
- [4] 柴田海翔, "複数の周波数成分に着目したプログラム理解の把握", 奈良工業高等専門学校情報工学科平成31年度卒業研究論文, 2020年3月.
- [5] Toyomi Ishida, Hidetake Uwano, "Time Series Analysis of Programmer's EEG for Debug State Classification," In the 5th Edition of the Programming Experience Workshop (PX 2019), April 2019.
- [6] Martin K.-C. Yeh, Dan Gopstein, Yu Yan, Yanyan Zhuang, "Detecting and comparing brain activity in short program comprehension using EEG", 2017 IEEE Frontiers in Education Conference (FIE), pp. 18-21 (2017)
- [7] W. Klimesch, "EEG- α rhythms and memory processes", International Journal of Psychophysiology, vol.26, pp.319-340 (1997).
- [8] Toyomi Ishida, Hidetake Uwano, and Yoshiharu Ikutani, "Combining Biometric Data with Focused Document Types Classifies a Success of Program Comprehension," In In the 28th IEEE/ACM International Conference on Program Comprehension (ICPC 2020), July 2020.
- [9] T. Fritz, A. Begel, S. C. Miller, S. Yigit-Elliott, M. Zger, "Using Psycho-Physiological Measures to Assess Task Difficulty in Software Development", In Proceedings of the International Conference on Software Engineering (ICSE), pp.402-413 (2014).
- [10] 村上郁也: "イラストレクチャー 認知神経科学", 株式会社オーム社 (2013).
- [11] 音茂龍司, 辻貞敏: "よくわかる脳波判読第3版": 金原出版株式会社 (2015).
- [12] 堀忠雄:"生理心理学-人間の行動を生理指標で測る": 培風館 (2008).
- [13] 坂本佑太, 吉田幸二, 宮地功: "簡易脳波計による学習時の思考と記憶の比較分析", マルチメディア, 分散, 協調とモバイル(DICOMO)2012シンポジウム論文集, pp.724-729(2012).