

システム創成工学専攻
情報システムコース

Department of Systems Innovation
Advanced Information System Course

令和6年度 専攻科特別研究論文

脳波を利用したプログラム理解の識別に
おける時系列情報の影響分析

Influence Analysis of the Time-series EEG Power
Spectrum for Program Comprehension Identification

指導教員名 上野 秀剛 准教授

論文提出者名 松山 鼓

独立行政法人 国立高等専門学校機構

奈良工業高等専門学校 専攻科

National Institute of Technology, Nara College

Faculty of Advanced Engineering

脳波を利用したプログラム理解の識別における時系列情報の影響分析

Influence Analysis of the Time-series EEG Power Spectrum for Program Comprehension Identification

松山 鼓

Matsuyama Tsuzumi

独立行政法人 国立高等専門学校機構

奈良工業高等専門学校 専攻科 システム創成工学専攻 情報システムコース

大和郡山市矢田町 22 番地 (〒 639-1080)

National Institute of Technology, Nara College, Faculty of Advanced Engineering

22 Yata-cho, Yamatokoriyama, Nara 639-1080, Japan

Abstract: Program comprehension is a critical task in software development, and enhancing its efficiency is essential. This study proposes a method to identify program comprehension states using EEG and investigates the impact of time-series data features on identification accuracy. Datasets with up to 15 divisions were created to compare 1) the presence or absence of overlapping time domains and 2) the effects of segmentation on time length. Additionally, 3) Dynamic Time Warping (DTW) was used to analyze the influence of individual differences in comprehension timing by calculating distances between events occurring at different speeds and timings. Results showed that overlapping time domains improve identification accuracy, with datasets segmented into 8 parts achieving the highest accuracy, consistent with prior findings. DTW analysis achieved an accuracy of 81.1 % using only α waves, comparable to methods using multiple frequency components. DTW demonstrated the ability to classify accurately despite individual differences in comprehension timing, performing well for simple tasks but proving less effective for complex tasks compared to conventional methods using multiple frequency components.

Keywords: EEG, Program Comprehension, Time-series division, DTW, Machine learning;

関連業績リスト

1. 松山 鼓, 上野秀剛, “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識別,” FIT2023 第 22 回情報科学技術フォーラム, 2023.

目次

1.	はじめに	1
2.	関連研究	4
3.	準備	6
3.1	脳波	6
3.2	DTW(Dynamic Time Warping)	8
4.	実験	10
4.1	実験環境	10
4.2	タスク	11
4.3	実験の手順	12
4.4	分析	13
5.	結果と考察	17
5.1	重複の有無と時間分割の増加	17
5.2	DTW によるクラスタリング	18
6.	おわりに	23
	参考文献	26

目次

3.1	国際式 10-20 電極法 [18]	6
3.2	DTW 例 [28]	9
4.1	実験概要	11
4.2	重複なしパワースペクトル算出法	14
4.3	重複ありパワースペクトル算出法	15
4.4	DTW 概要	16
5.1	重複ありの精度平均	18
5.2	重複なしの精度平均	18
5.3	階層クラスタリング	19
5.4	各時系列データのグラフ	22

表目次

3.1	周波数成分 (5 種)	7
3.2	周波数成分 (8 種) [12,13]	8
4.1	タスク一覧	12
5.1	DTW による識別結果	19
5.2	被験者ごとの識別精度	20
5.3	タスクごとの識別精度	21

1. はじめに

プログラム理解はバグの修正やコードの再利用や変更などソフトウェア開発における様々な工程で行われる重要な作業である。開発者は開発に費やす時間の約 70% をソースコードの理解に費やしており [1]、その効率・品質を向上させることはプログラム工学の重要な研究テーマの 1 つである。プログラム理解は人の活動であり、ロジカルシンキングや数値処理、シンボル操作、言語処理といった様々な認知機能が用いられる [2]。そのため、開発者のプログラム理解中の様子を脳活動や視線計測などの生体情報に着目して計測し、理解の正否や効率、精神作業の遂行に伴う負荷であるメンタルワークロード (MWL) との関係进行分析する研究が多数行われており、中でも脳波は計測のしやすさや被験者への負荷の低さなどから広く用いられている [3–10]。

脳波は脳で生じる電氣的活動を計測したものであり、国際脳波学会によって周波数帯域ごとに名称が付けられている。脳波の波形特徴や高速フーリエ変換 (FFT) で得られる周波数成分のパワースペクトルは被験者の状態によって変化するため、様々な研究において指標として用いられる。例えば、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 波が多く出現する [11]。プログラム理解の分野においても周波数成分と理解状態の関係を分析した研究が多い。Yeh らはプログラム理解中に困難を感じている被験者の α 波と β 波に有意差が見られることを報告している [9]。Lee らはプログラム理解タスクにおいて初心者と熟練者の間で β 波と γ 波に有意な差を計測している [4]。

著者のこれまでの研究では、プログラム理解タスク中の脳波を一定時間ごとに分割し、周波数成分の時系列変化を特徴量とした機械学習によってプログラム理解の有無を識別できるか評価している。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、時間分解能が高い。プログラム理解の作業中は理解の段階に応じてリアルタイムで思考が変化する為、作業中の脳波を時間で分割して得られる周波数成分はプログラム理解の状態を識別する有用な特徴であると考えられる。

実験ではプログラム理解タスクを行う被験者を対象に計測した脳波を異なる区間数に分

割した複数のデータセットを作成し、区間ごとの周波数成分を求め、データセット間で識別精度を比較している。また、脳波の周波数帯域には一部の帯域を細かく分けた定義も存在し、ソフトウェア工学を含む複数の分野で指標として用いられている [6,7,12,13]。

分析では、プログラム理解を行う開発者の脳波に対して、8種 ($\delta, \theta, L\alpha, H\alpha, L\beta, H\beta, L\gamma, M\gamma$), 5種 ($\delta, \theta, \alpha, \beta, \gamma$) の周波数成分を特徴量とした機械学習を実施してきた。プログラム理解タスク中のリアルタイムな脳波の変化を捉えるために脳波データを複数の時間領域に分割し、各時間帯から FFT でパワースペクトルを算出する。データ分割時には、連続する2つの時間領域が50%重なるように、また、データ長が等しくなるように分割した。分割数は1~10に変化させてデータセットを作成した。その後、全タスクの全時間帯の周波数成分の平均値を用いて正規化した周波数成分を特徴量としてランダムフォレストによる分類を行い、プログラム理解の有無を識別した。分析の結果、5種類の周波数成分を用い、時間分割数を8とした場合に、識別精度が84.1%と最大になった。また、予測にもっとも重要な周波数成分が α 波であることが確認された。一方で、時間領域の重複をなくすことや、時間分割数を増加することで、理解の進行に伴うリアルタイムな脳波の変化をより詳細に捉えることができ、理解状態に発生する最も特徴的なデータのみを抽出して識別精度が向上する可能性がある。また、プログラム理解のタイミングには個人差が存在し、理解タイミングが通常と異なる場合には識別の精度が低下する懸念も存在する。

本研究では、識別精度への影響を明らかにするために、1) 時間領域の重複の有無と、2) 時間分割数を分析対象データの時間長に対して妥当な分割数の最大値である15まで変化させたデータセットを作成し、予測精度を比較する。また、理解タイミングの個人差の影響を明らかにするために3) 時間軸を伸縮することで異なる速度やタイミングで発生するイベント間の距離を算出できるDTWを用いたクラスタリングを行う。分割数を増加させることで、脳波のリアルタイムな変化をより詳細に捉え、理解に関連する特徴的なデータを抽出することを目指す。さらに、50%重複させた分割では、隣接する時間帯の情報を統合し、一貫性のある特徴量を生成することで識別精度の向上が期待される。脳波は常に変化を続けているため、特定の瞬間でのみ区切ってしまうと、重要な変化を見逃してしまう可能性がある。そこで、隣接する時間帯の情報を活用することで、各時間帯の特徴を安定させることができ、高精度が得られると考えられる。一方、重複させない分割では、冗長な情報を排除し、重要な特徴を際立たせることが可能となる。重要な特徴が一部の瞬間にのみ大きく発生している場合、その瞬間の特徴のみを抽出することができ、高精度に繋がる可能性があると考えられる。このように、分割数の増加や重複の有無が識別精度に与

える影響を検証する。また、 α 波のみを対象に DTW を用いた解析を行うことで、プログラム理解に関連する脳波の時間依存性をより効果的に評価できると考える。DTW は、時間軸を非線形に伸縮して時系列データ間の類似性を測定できるため、理解のタイミングに個人差がある場合でも適切に比較が可能である。これにより、従来の手法では捉えきれなかった個人差を補正し、識別精度の向上が期待される。

以下、2 章では関連研究について説明し、3 章では脳波,DTW 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、脳波計測装置の高性能化・低価格化に伴い、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 functional Magnetic Resonance Imaging (fMRI) を用いてプログラム理解時の脳活動を観測した [14]。fMRI は脳の領域ごとの血流変化を測定することでどの領域が活性化するかリアルタイムに測定する。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 と比較して計測時の姿勢や体動の制限が少なく、より実際の環境に近い状態での計測が可能な脳波計測装置を用いた研究も多数行われている。Yeh らは C/C++ のソースコードを対象としたプログラム理解時の脳波を計測し、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実験した [9]。実験の結果、高難易度のソースコードを理解するときに $High\alpha$ 波と θ 波のパワースペクトルが増加した。Lee らは Java の小規模なソースコードを用いたプログラム理解タスクにおいて初心者と熟練者を計測し、前頭葉で β 波と γ 波に有意な差が見られた [4]。Fritz らはプログラム理解タスク中の脳波、視線移動、皮膚電気活動を計測し、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識別した [15]。結果として、タスクの難易度（易しい/難しい）を 65.0% の精度、64.6% の再現率で識別し、被験者が初めて行うタスクの場合 84.4% の精度、69.8% の再現率で識別できた。石田らはプログラムの動作を理解するタスクとバグの有無を判断するタスクを含む実験を行っている。実験の結果、プログラムの動作を理解するタスクにおいて、正しく理解できた被験者の α 波が有意に大きい一方で、バグの有無を判断するタスクにおいては、正しく判断できた被験者の β 波が有意に大きく、有意差のある周波数に差が見られた [16]。

プログラム理解を行っている作業者の思考はプログラムに対する理解状況によってリアルタイムに変化する。例えば、読み始めた直後はクラスやメソッド、ブロックがどのような順で呼ばれているのか、プログラムの全体像を大まかに理解する一方で、後半には仮の値を代入するなどしながら、個々の処理について詳細な理解を進める。そのため、理解中の脳波を時系列に分割し、それぞれの周波数成分の強さを特徴量とすることで、理解状態の識別に有用と考えられている。上記の理由から、石田と上野は異なる研究 [17] においてプログラム理解タスクにおけるタスク開始直後とタスク終了時の変化を分析している。脳波が計測された時間に着目している点で本研究と類似しているが、石田の研究ではタスク中の脳波は分析対象になっておらず、また、分析の対象も α 波、 β 波のみである。本研究ではタスク中に計測されたすべての脳波を抽出対象としており、また、5 種類の周波数成分と 8 種類の周波数成分それぞれを用いて、機械学習による分類精度を評価する点で違いがある。

3. 準備

3.1 脳波

3.1.1 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，頭皮上に装着した電極から計測される．電極の配置は国際式 10-20 電極法に則って行う (3.1)．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．

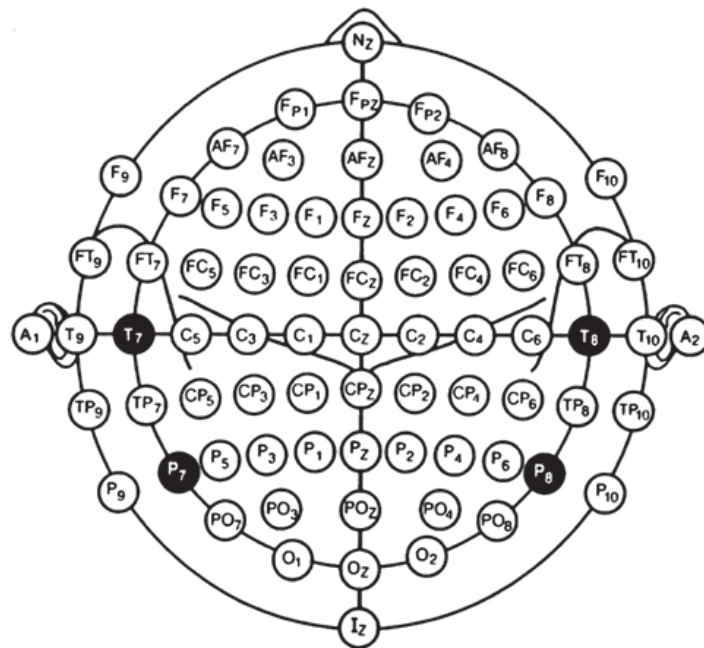


図 3.1 国際式 10-20 電極法 [18]

表 3.1 周波数成分 (5 種)

種類	周波数帯域 [Hz]
δ 波	0.5~4
θ 波	4~8
α 波	8~14
β 波	14~30
γ 波	30 以上

脳波の計測方法には基準電極導出法と双極導出法の 2 種類がある [19]. 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. 探査電極は国際式 10-20 電極法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, 脳電位の電場内に置いた 2 つの電極の電位差として脳電位を測定する. 2 つの電極に共通する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.

3.1.2 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. 周波数成分のパワースペクトルを抽出するために FFT 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 [11, 19] である 5 種類の周波数成分を表 3.1 に示す.

δ 波は睡眠状態にある時に出現し, θ 波は感情状態と関連 [20] がある. 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. また, 音声の聞き取りのような認知処理と関連がある [21]. β 波は緊張や不快な感情を抱いているなどストレスが強い時や論理的思考時に出現するほか, 視覚刺激 [22] と関連する. γ 波は, 不安で興奮している時に出現しやすい. 周波数成分のパワースペクトルは様々な脳活動との関係が分析されており [23], 運動時の認知負荷 [24], 空間認知時の MWL [25], 記憶能力 [26], 選択的注意の有無 [27] などとの関係が明らかになっている. プログラム理解の分野においても様々なタスクと周波数成分の関係が分析されており, 文献 [5] ではプログラムの動作を想起した際の α 波と θ 波が, 文献 [10] ではプログラム理解タスク中の α 波, β 波, γ 波, θ 波について実験により分析している. 一部の周波数帯域については更に細かくした分類も存

表 3.2 周波数成分 (8 種) [12,13]

種類	周波数帯域 [Hz]
δ 波	0.5~2.75
θ 波	3.5~6.75
Low α 波	7.5~9.25
High α 波	10~11.75
Low β 波	13~16.75
High β 波	18~29.75
Low γ 波	31~39.75
Mid γ 波	41~49.75

在しており，複数の研究 [6,7,12,13] で分析の指標として用いられている．本研究では，5 種類の分類に加え，表 3.2 に示す 8 種類の分類を特徴量として時系列の変化を用いた機械学習を行う．

3.2 DTW(Dynamic Time Warping)

DTW(Dynamic Time Warping) とは，異なる長さや速度で進行する時系列データ間の類似性を測定する手法である．図 3.2 に DTW の例を示す．DTW は，2 つの時系列間で時間軸を非線形に調整し，最も類似する形状になるようにデータ間の対応付けを行う．図 3.2 に示す二つの時系列データを結ぶ線がデータ間の対応づけされた点となる．これらの対応するデータの距離を算出することで，時系列データ間の類似度が求められる．これにより，速度やタイミングの違いを無視して，パターンの類似性を評価できることが特徴である．DTW は，時系列データ $Q = \{q_1, q_2, \dots, q_n\}$ と $C = \{c_1, c_2, \dots, c_m\}$ の間で，対応する点の累積距離を最小化する経路を探索することで実現される．この DTW 距離は以下のように定義される．

$$D(i, j) = \text{dist}(q_i, c_j) + \min\{D(i-1, j-1), D(i-1, j), D(i, j-1)\} \quad (3.1)$$

$\text{dist}(q_i, c_j)$ は点 q_i と c_j の間の距離（ユークリッド距離）を表す．この式を用いて累積距離行列 D を計算し，最小経路を求めることで，時系列間の類似性を評価する．DTW の

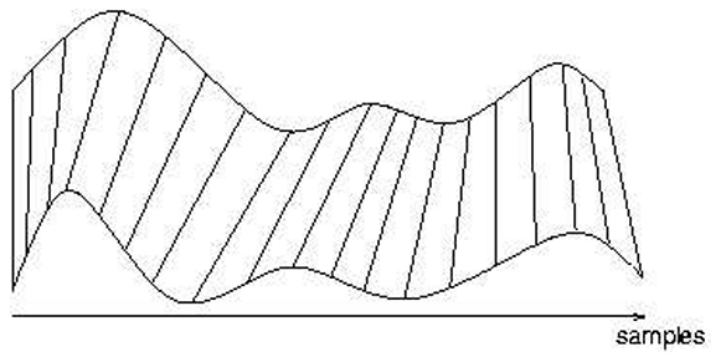


図 3.2 DTW 例 [28]

大きな利点は、時間的な変動や速度の違いを考慮できる点にある。

4. 実験

本研究では先行研究 [16] の実験で計測したプログラム理解タスク中の脳波データを用いた分析を行う。本実験を行うにあたり、事前に奈良高専の倫理審査委員会の承認（奈高専倫 21-40）を得て実施している。被験者には実験の前に目的・方法・予測される問題等について書面で説明し、同意を得た上で実験を行った。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 of 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履修済みである。

4.1 実験環境

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。体動によって生じる電気的なノイズであ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。本計測装置の計測周期は 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。本計測装置の制御とデータ取得には BioTrace+ Software を用いる。

脳波を計測する電極の配置は国際式 10-20 電極法に基づき、基準電極導出法を用いる。探査電極を後頭部 (Pz)、基準電極を左耳 (A1)、グラウンド電極を右耳 (A2) に配置した。後頭部 (Pz) は筋電位など脳波以外の生体現象のアーチファクトが入りにくく、プログラム理解を対象とした研究においてもよく計測されている [4,7,8]。

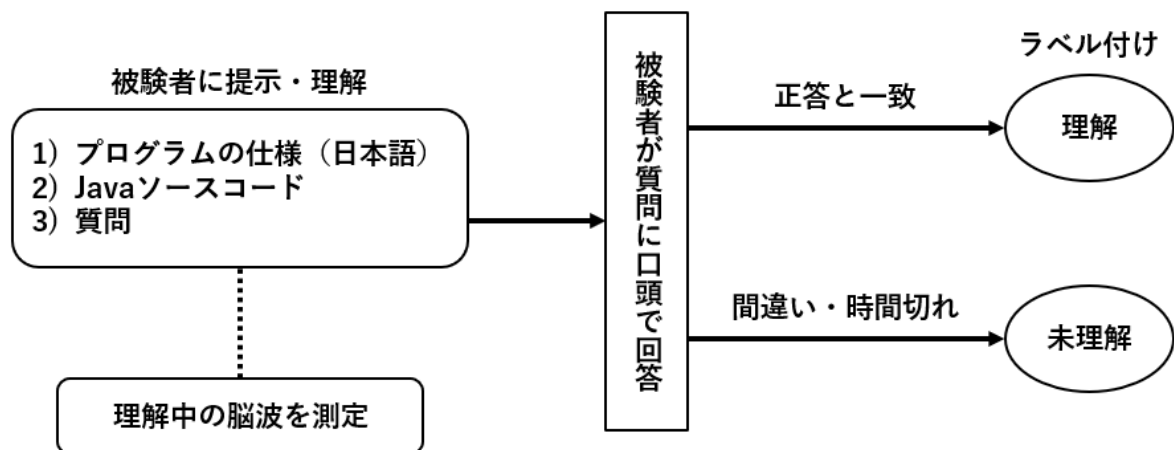


図 4.1 実験概要

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。概要を図 4.1 に示す。被験者に 1) 日本語で記述されたプログラムの仕様と、2) 対応する Java のソースコード、3) 質問の 3 点を 1 組として提示する。被験者は仕様とソースコードを元にプログラムの動作を理解し、質問に口頭で回答する。質問は日本語で記述されており（2 回目のループの時の 6 行目の x の値は？）、回答（ x の値は 7）が事前に用意した正答と一致していれば動作を理解している（「理解」とラベル）とみなす。回答が正答と一致していない、または、制限時間以内に回答できなければ理解していない（「未理解」とラベル）とみなす。回答の正誤は被験者には伝えない。

本実験では、理解と未理解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（2 分 30 秒）は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

表 4.1 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		順列を会列挙する
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

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 4.1 にタスク一覧を示す。

4.3 実験の手順

実験の順序を以下に示す。

1. 実験の説明
 実験概要、脳波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. 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. 回答
 - 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. 全タスクの実施
 - 16タスクが完了するまで手順4~5を繰り返す。

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、各被験者のタスクごとの脳波データを複数の時間領域に分割し、パワースペクトルを求める。パワースペクトル算出の際には、1) 重複なしで等分割したものと、2) 分割後のデータ長が等しくなるよう重複(50%)させたものの、2種類のデータを用意する。図4.2、図4.3にそれぞれのパワースペクトル算出法を示す。

重複無しの算出では、被験者 s のタスク t で計測した脳波 $BW_{s,t}$ を n 等分し、分割後の個々の脳波 $BW_{s,t,d}$ に対してFFTで周波数成分ごとのパワースペクトルを求める(図4.2)。重複ありの算出では、被験者 s のタスク t で計測した脳波 $BW_{s,t}$ を n 個に分割する際に、 x 個目の分割脳波 $BW_{s,t,x}$ が $x-1$ 個目の分割後脳波 $BW_{s,t,x-1}$ のと50%重複するよう時間長を求める。分割後の個々の脳波 $BW_{s,t,d}$ に対してFFTで周波数成分ごとのパワースペクトルを求める。脳波の個人差を考慮し、 n 等分した被験者 s 、タスク t の時間帯 $d(1, \dots, n)$ における周波数成分 $Freq$ のパワースペクトルを、被験者 s の全タスクにおけるすべての時間帯の全周波数成分の平均で正規化する。

$$Norm.Freq_{s,t,d} = \frac{Freq_{s,t,d}}{Ave(\delta_{s,1,1}, \dots, Mid \gamma_{s,1,n}, \delta_{s,16,1}, \dots, Mid \gamma_{s,16,n})} \quad (4.1)$$

正規化した各時間帯の周波数成分をメトリクスとしてRandom forestによって機械学習させる。正解集合として実験時に集計した理解/未理解を用いる。Leave one out法を

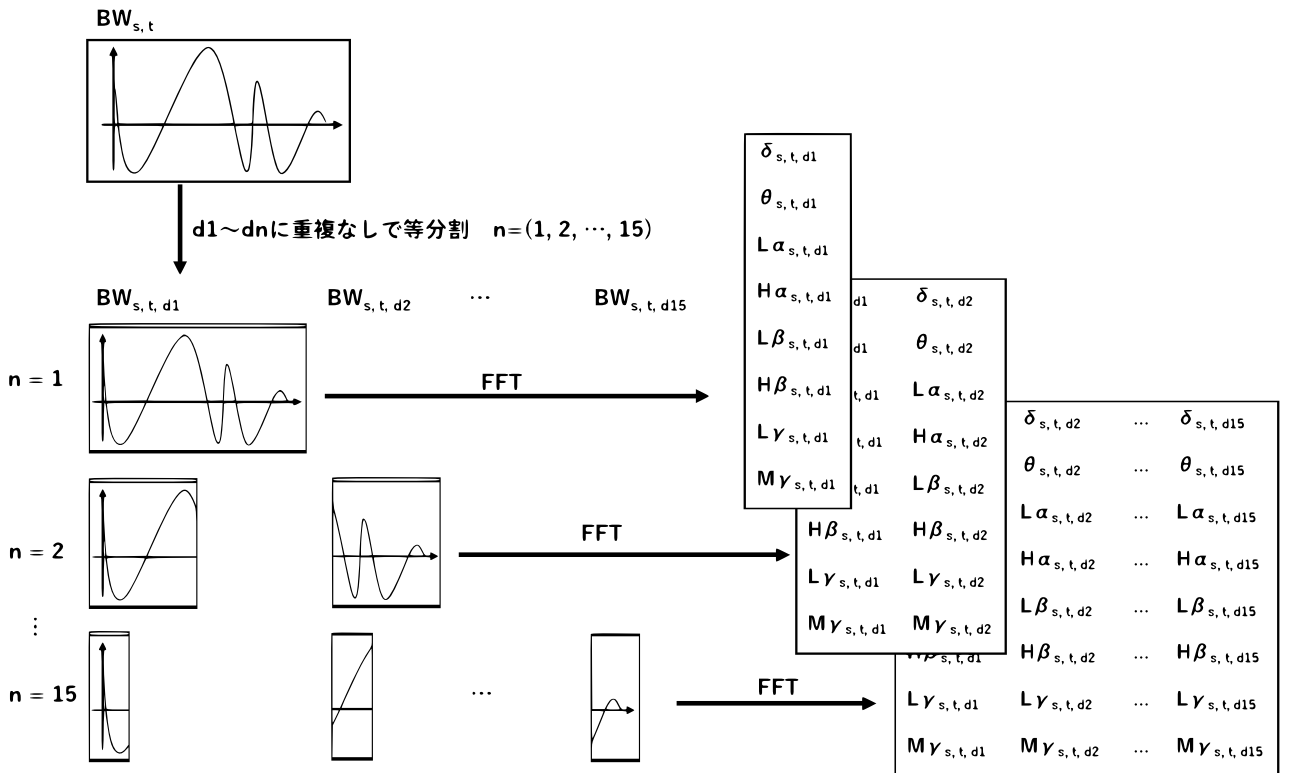


図 4.2 重複なしパワースペクトル算出法

用いてすべてのタスクデータに対して識別を行い精度を算出する。Random forest のパラメータ設定は決定木モデル数 $ntree$ を 500, 使用メトリクス数 $mtry$ を予備実験で行ったモデルチューニングにおいてエラー率の小さかった 1 から 16 までの値を用いて分析する。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。パワースペクトルを算出する際に用いる周波数成分の識別方法の違いが識別精度に与える影響を評価するために, 5 種類 ($\delta, \theta, \alpha, \beta, \gamma$) の周波数成分と, α 波, β 波を Low, High に, γ 波を Low, mid にそれぞれ識別した 8 種類 ($\delta, \theta, L\alpha, H\alpha, L\beta, H\beta, L\gamma, M\gamma$) の周波数成分を用いる。また, 時間の分割数 n についても分割数による識別精度の差を評価するため, 1 分割 (分割無し) から 15 分割まで変化したデータセットを作成し, 識別精度を比較する。特に, これまでの研究で最も高精度が確認できた 8 分割時と 11 分割以降で有意差が発生するか確認する。

また, 時系列データを詳細に分析できるよう 15 分割時の α 波のデータに対して DTW を用いた時系列解析を実施し, プログラム理解状態を識別するためのクラスタリングを行

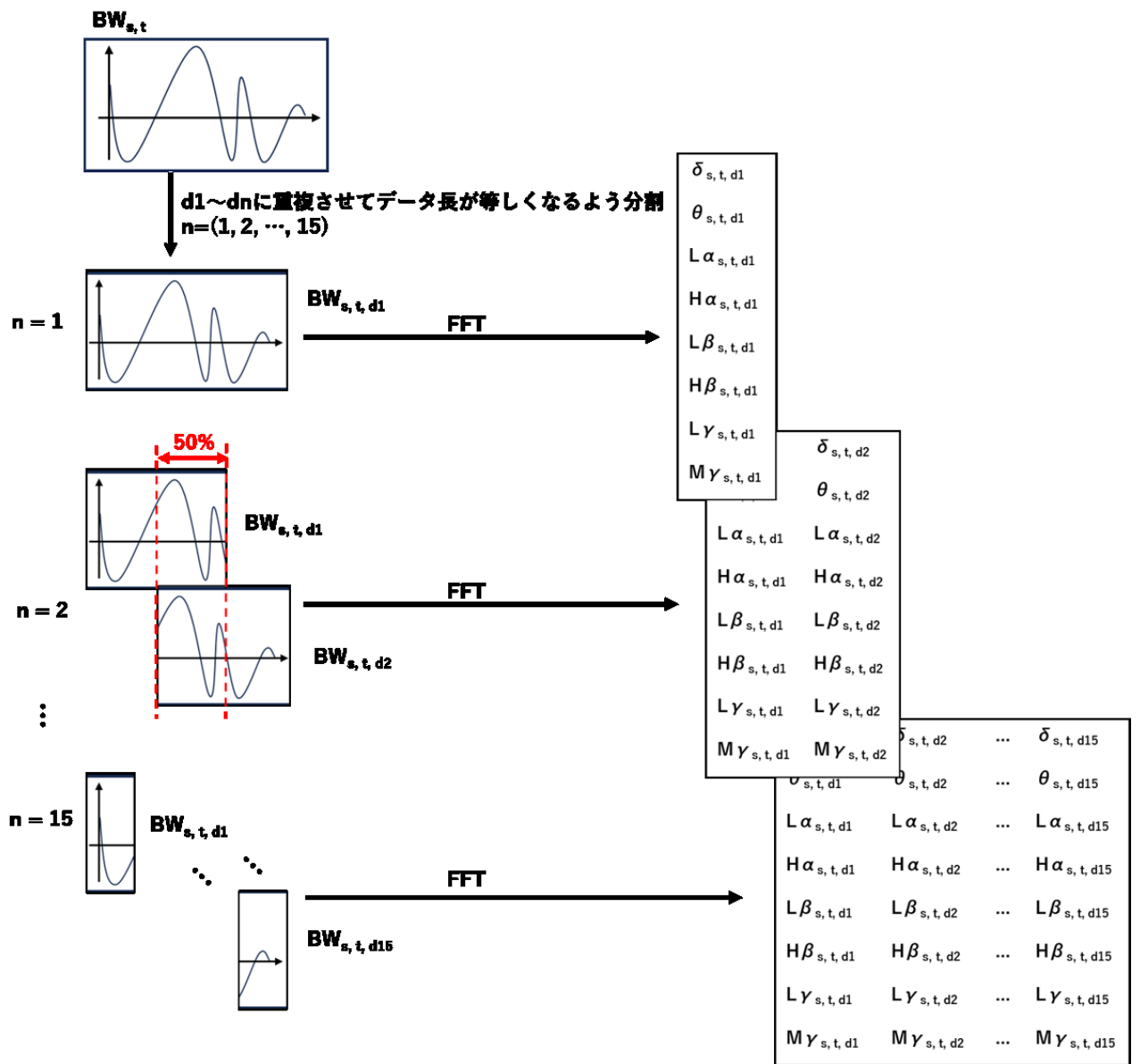


図 4.3 重複ありパワースペクトル算出法

図 4.4 に DTW 使用の流れを示す. 各被験者, 各タスクの 15 分割時のデータをもとに, それぞれの時系列データを作成する. 時系列データは, 各被験者及びタスクごとに個別で構成し, DTW への入力とする. DTW では, 1) 各被験者およびタスクごとの時系列データを比較して距離を算出し, 2) 得られた距離を基に距離行列を作成, 3) 距離行列を基にク

ラスタ数を 2 に設定し階層クラスタリングを行い, 4) 各クラスタに理解状態を示すラベル Yes/No を割り当てる.

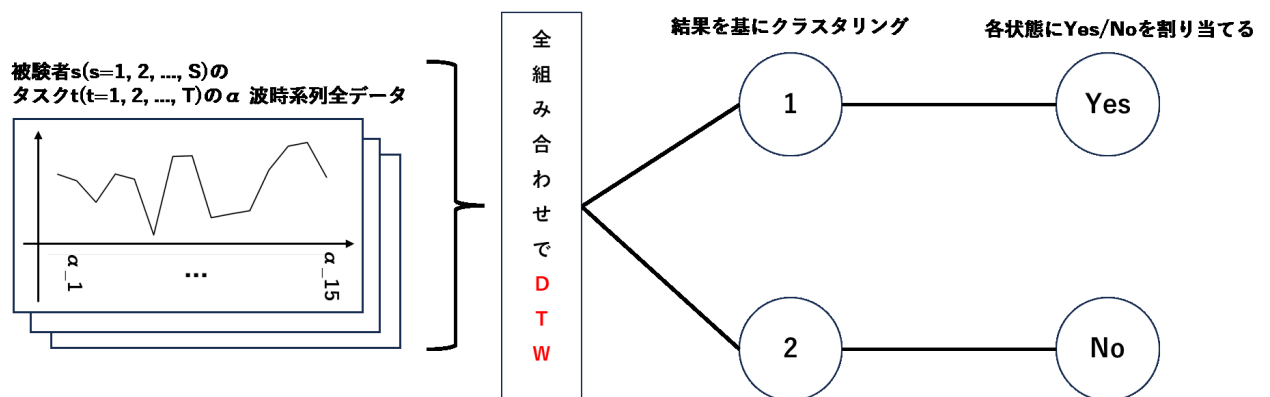


図 4.4 DTW 概要

5. 結果と考察

実験の結果、5人×16タスクの合計80件のデータを取得した。タスク中に電極が計測部から外れたため正常に計測できなかった6件を除いた74件（理解46件、未理解28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 $n=1\sim 15$ の15個と、脳波を5種類の周波数成分とした $n=1\sim 15$ の15個の計30個について評価する。

5.1 重複の有無と時間分割の増加

周波数成分8種と5種それぞれにおける、重複ありの分割数ごとの識別精度の平均を図5.1に、重複なしの分割数ごとの識別精度の平均を図5.2に示す。横軸は分割数、縦軸は識別精度の平均値、線の種類は周波数成分の種類数を表す。重複ありと重複なしの最も精度が高かった精度を比較する。重複ありでは最高0.845（5種、8分割）、重複なし0.812（5種、5分割）となった。15分割まで増加させ、重複を比較した予測においても、以前の研究と同様に5種、8種が最適であることが確認できる。

分割数の増加については、重複ありでは11～14分割で精度が8分割と比較して低下しているものの、15分割では再び精度が向上する傾向が見られる。一方、重複なしでは5分割時に最高精度を記録したものの、11分割以降においても精度の向上が見られず、むしろ変化が小さい結果となった。

これらの結果から、重複ありの場合は隣接する時間領域間の情報を共有することで、時間的な連続性のある特徴量が抽出され、精度向上に繋がっていると考えられる。また、11分割以降で精度が低下している要因は、分割数を増やすことでデータ量が減り、時間的な連続性のある一部の重要な特徴が抽出できなくなったことだと考えられる。冗長な情報を排除し、重要な特徴を際立たせるといった期待通りの結果は得られなかった。これらは情報量と特徴量のバランスが精度向上に重要であることが示している。8分割が、現タスクの

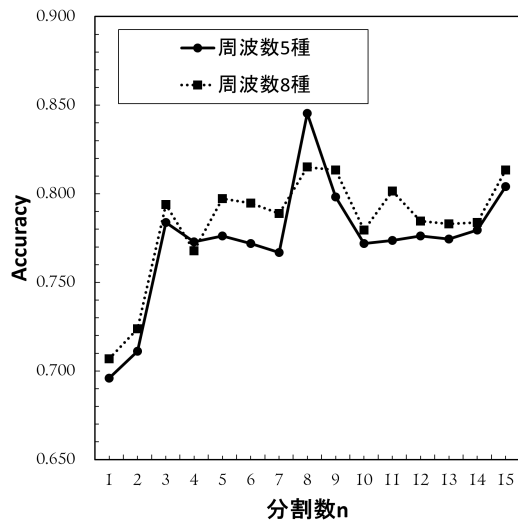


図 5.1 重複ありの精度平均

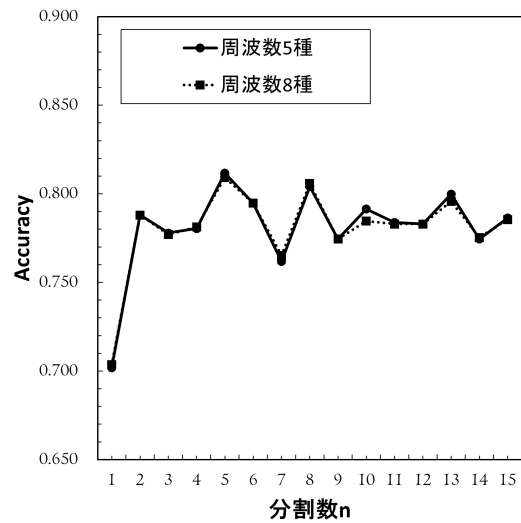


図 5.2 重複なしの精度平均

時系列情報から得られる各データが持つ情報に対して、理解時の特徴を最もバランス良く抽出できるパラメータであることが考えられる。一方で、8分割時の結果には及ばなかったものの、重複ありで15分割時に精度がわずかに向上したことから、プログラム理解時に発生する特徴には、予測精度には大きく寄与しないが尖鋭的な特徴が含まれている可能性が示唆された。15分割時の精度向上や周波数成分の選択による効果については、今後の分析でより詳細に検討する必要がある。周波数成分8種と5種の比較では、分割数11以降も全体的に精度に大きな差は見られなかった。上記結果から、重複ありの5種、8分割が最適な分割設定であり、適切な分割数と重複設定が識別精度を向上させる重要な要因であることが示された。

5.2 DTW によるクラスタリング

図 5.3 に DTW による階層クラスタリングの結果を示す。横軸が被験者、タスク、縦軸が DTW の距離を示している。左右2つのクラスターが離れた距離に存在しており、図左側に未理解、図右側に理解したタスクが集中した。一方で、一部のタスクでは理解／未理解が異なる群に分類されている（図の色つきのタスク）。主に理解が属するクラスターを理解クラス、主に未理解が属するクラスターを未理解クラスとした場合の混同行列を表 5.1 に示す。60/74(81.1%) が正しくクラスタリングされており、この精度は5種類の周波数成分

表 5.2 被験者ごとの識別精度

被験者	DTW	Random forest
Sbj.A	85.7%	92.9%
Sbj.B	66.7%	73.3%
Sbj.C	71.4%	78.6%
Sbj.D	87.5%	93.8%
Sbj.E	93.3%	100.0%

Random forest で 73.3% と差が見られる。タスクごとの精度 (表 5.3) に着目すると難易度が easy のタスク (タスク 1~タスク 8) はいずれも DTW と Random forest で同じ精度だった一方で、難易度が difficult のタスクのうち、タスク 11,13,14 で DTW の精度が低く、最大で 50% の差が見られた。この結果は easy タスクは DTW が用いている α 波のみのデータで複数の周波数成分を用いた従来研究と同等の結果が得られる一方、difficult タスクの識別には α 波のみでは不十分であり、 β 波など他の周波数成分が必要である可能性を示唆する。今後、 β 波などの他の周波数成分を用いた時の DTW の精度を評価すると同時に、複数の周波数成分を同時に用いた分析方法を提案することでより高い精度で識別が行える可能性がある。

図 5.4 に α 波成分の時系列変化を理解/未理解で色分けしたものを示す。理解したタスク (赤色) は、未理解のタスク (青色) と比較して値が全体を通じて高く、増減しながら推移している。未理解のタスクは一時的に高い値になる場合が見られるが、いずれも低い値で安定している。また、理解したタスクの山が一つに重なっていないことから、理解タイミングが別々だということが確認できる。

このような特徴のでるタイミングの違う時系列データを理解/未理解に高精度で分類できていることは、DTW を用いることで時系列データを柔軟に比較できたからだと考えられる。また、デンドログラムによる可視化を通じて、分類結果の妥当性や各クラスタの特徴を容易に検証することができた。一部のデータにおいては、「Yes」と「No」の境界が曖昧であり、識別が困難なケースが存在することが確認された。これらのデータについては、 β 波や γ 波など他の周波数成分を補足的に利用することで、さらなる精度向上が期待できる。また、時系列データの特定の時間帯が識別精度に与える影響を詳細に分析し、特徴的なデータを抽出する手法を検討する必要がある。

著者のこれまでの研究では、周波数成分が 5 種または 8 種の場合において、個別の成分

表 5.3 タスクごとの識別精度

タスク	難易度	DTW	Random forest
1	easy	100.0%	100.0%
2		80.0%	80.0%
3		80.0%	80.0%
4		100.0%	100.0%
5		80.0%	80.0%
6		75.0%	75.0%
7		100.0%	100.0%
8		80.0%	80.0%
9	difficult	100.0%	100.0%
10		80.0%	80.0%
11		50.0%	100.0%
12		80.0%	80.0%
13		80.0%	100.0%
14		20.0%	60.0%
15		100.0%	100.0%
16		100.0%	100.0%

が識別精度に与える影響が異なることが確認されている。 α 波が半分以上の重要度を持っており、それ以外の周波数成分が β , θ , γ , δ という順番で重要度を持っていた。本研究での DTW による識別でも α 波のみで以前の研究ほぼ同等の精度で識別できている。特に、easy タスクでの識別が高精度で行えている。一方で、difficult タスクでは識別精度が低いことから、簡単なタスクでは α 波が、複雑なタスクではそれ以外の成分の重要度が増加していると考えられる。これらの結果は、構造が簡単なタスクでは α 波による DTW 識別、複雑なタスクでは複数の周波数成分を用いた識別が有効である可能性を示唆している。

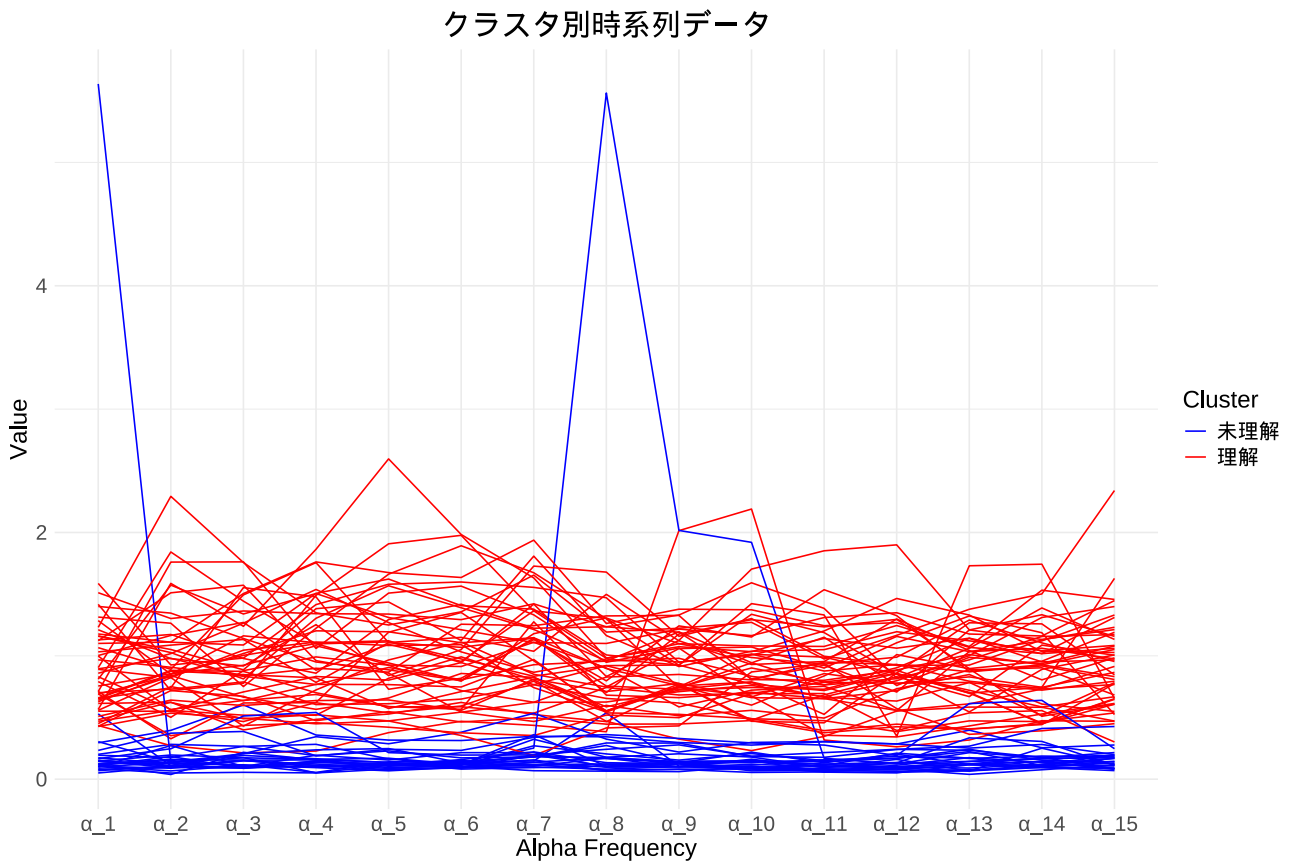


図 5.4 各時系列データのグラフ

6. おわりに

本研究はプログラム理解タスク中の脳波から理解／未理解を識別する際の精度への影響を明らかにするために、1) 時間領域の重複の有無と、2) 時間分割数を分析対象データの時間長に対して妥当な分割数の最大値である 15 まで変化させたデータセットを作成し、予測精度を比較した。また、理解タイミングの個人差の影響を明らかにするために 3) 時間軸を伸縮することで異なる速度やタイミングで発生するイベント間の距離を算出できる DTW を用いたクラスタリングを行った。

著者の従来研究である周波数成分を特徴量とする Random forest による機械学習のデータに対して時間領域の重複あり (50 %) と重複なしの比較を実施した結果、従来研究の結果と同様に 5 種の周波数成分を用いた 8 分割時に識別精度が高くなることが確認できた。また、重複ありの分割設定が識別精度を向上させる重要な要因であることが確認された。冗長な情報を排除し、重要な特徴を際立たせるという期待通りの結果は得られなかったが、改めて重複ありの 5 種、8 分割が最適な分割設定であり、適切な分割数と重複設定が識別精度を向上させる重要な要因であることが示された。

従来の結果にて、周波数成分の中で重要度の高い α 波を採用し、DTW を用いた時系列データのクラスタリングを実施した。2 値にクラスタリングした後に、実際の理解/未理解と比較して、Yes/No (理解の有無) のラベル付けを行った。ラベル付けの結果と実際の理解/未理解が一致する割合を算出した結果、 α 波のみを対象としながらも識別精度 81.1 % を達成した。DTW を用いたクラスタリングでは α 波のみを用いていることから、複数の周波数成分を用いることで従来手法より高い精度で識別できる可能性が示された。一方で、DTW を多次元データに対して適用させることが困難であり、現状単一の周波数成分のみでしか識別が難しいことは課題として挙げられる。各被験者、各タスクの識別結果の内訳を確認すると、easy タスクに関しては、従来結果と同程度で高精度で識別が行えた。このことから、DTW を用いることで簡単なタスクに関しては、従来研究よりもはるかに少ない α 波のみのデータで理解の有無を識別できる事が示された。一方で、difficult

タスクに関しては、従来結果を大きく下回る結果となった。これは、構造が複雑なタスクでは、 α 波以外の周波数成分の情報が重要になることを示唆している。著者のこれまでの研究では、周波数成分 5 種を 8 分割したデータに Randon Forest を用いることで得られている最大 83.8 % の識別精度はすでに一定の実用性を示している。DTW の全体的な精度では、ほぼ同程度の精度が得られたものの、数値だけでみれば Random Forest の結果には劣ってしまう。しかし、本研究によって構造が簡単な場合では α 波を用いた DTW が有効となることなど、より高い精度で識別するための知見が得られたことで、理解が進んでいない学生を漏らさずに支援することが可能になり、教育効果を高めることができるようになると思われる。

結果から、プログラム理解の識別には、時間領域は重複させること、分割数は多くさせすぎず、本研究データでは 8 分割が有効であることが確認できた。また、DTW を用いることで特徴が現れるタイミングに個人差がある場合でも、時系列データを柔軟に比較できるため、分類の精度向上に寄与したと考えられる。今後は、 β 波などの他の周波数成分を用いた DTW 分析や、他の分割数を用いた DTW の分析を行うことで、課題解決する手法を模索する必要がある。また、そのために、DTW を用いて複数の周波数成分の時系列データを多次元的に分析する手法も検討する必要がある。上記課題に取り組むことで、プログラム理解の支援手法としての実用性を高め、脳波データ解析の新たな応用可能性を示すことが期待される。

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。査読教員である松尾賢一教授には，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] Minelli, R., Mocci, A., and Lanza, M.: I know what you did last summer: an investigation of how developers spend their time. In Proc. the 2015 IEEE 23rd Int'l Conf. on Program Comprehension. IEEE Press, pp.25–35 (2015).
- [2] Castelhana, J., Duarte, I. C., Ferreira, C., et al.: The role of the insula in intuitive expert bug detection in computer code: an fmri study, Brain imaging and behavior, pp.1–15 (2018).
- [3] Kosti, M., Georgiadis, K., Adamos, D., Laskaris, N., Spinellis, D., and Angelis, L.: Towards an Affordable Brain Computer Interface for the Assessment of Programmers' Mental Workload. Int. J. Human-Computer Studies, Vol.115, pp.52–66 (2018).
- [4] Lee, S., Matteson, A., Hooshyar, D., Kim, S., Jung, J., Nam, G., and Lim, H.: Comparing Programming Language Comprehension between Novice and Expert Programmers Using EEG Analysis. In Proc. Int' l Conf. on Bioinformatics and Bioengineering (BIBE). IEEE, pp.350–355 (2016).
- [5] Crk, I., and Kluthe, T.: Toward Using Alpha and Theta BrainWaves to Quantify Programmer Expertise. In Int' l Conf. Engineering in Medicine and Biology Society. IEEE, pp.5373–5376 (2014).
- [6] Lin, Y.-T., Liao, Y.-Z., Hu, X., and Wu, C.-C.: EEG Activities During Program Comprehension: An Exploration of Cognition. IEEE Access 9, pp.120407–120421 (2021).
- [7] Medeiros, J., Couceiro, R., Castelhana, J., Castelo Branco, M., Duarte, G., Duarte, C., Durães, J., Madeira, H., Carvalho, P., and Teixeira, C.: Software Code Complexity Assessment Using EEG Features. In Proc. 41th Int' l Conf. Engineering

- in Medicine and Biology Society. IEEE, pp.1413–1416 (2019).
- [8] Medeiros, J., Couceiro, R., Duarte, G., Durães, J., Castelhana, J., Duarte, C., Castelo-Branco, M., Madeira, H., Carvalho, P., and Teixeira, C.: Can EEG Be Adopted as a Neuroscience Reference for Assessing Software Programmers' Cognitive Load? *Sensors* 21, No.7, p.2338 (2021).
 - [9] Yeh, M., Gopstein, D., Yan, Y., and Zhuang, Y.: Detecting and Comparing Brain Activity in Short Program Comprehension Using EEG. In *Frontiers in Educ. Conf. IEEE*, pp.1–5 (2017).
 - [10] Peitek, N., Bergum, A., Rekrut, M., Mucke, J., Nadig, M., Parnin, C., Siegmund, J., and Apel, S.: Correlates of Programmer Efficacy and Their Link to Experience: A Combined EEG and Eye-Tracking Study. In *Proc. the 30th ACM Joint European Software Engineering Conf. and Symposium on the Foundations of Software Engineering (ESEC/FSE 2022)*. Association for Computing Machinery, pp.120–131 (2022).
 - [11] 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, p.344, 北大路書房, 京都 (1998).
 - [12] Jenke, R., Peer, A., and Buss, M.: Feature extraction and selection for emotion recognition from EEG. *IEEE Trans. Affect. Comput.*, Vol.5, No.3, pp.327–339 (2014).
 - [13] Memar, A. H., and Esfahani, E. T.: Objective Assessment of Human Workload in Physical Human-robot Cooperation Using Brain Monitoring. *J. Hum.-Robot Interact.* Vol.9, No.13, pp.1–21 (2019).
 - [14] Siegmund, J., Kanster, C., Apel, S., Parnin, C., Bethmann, A., Leich, T., Saake, G., and Brechmann, A.: Understanding Understanding Source Code with Functional Magnetic Resonance Imaging. *Proc. Int'l Conf. on Software Engineering (ICSE)*, pp.378–389 (2014).
 - [15] Fritz, T., Begel, A., Müller, S. C., Yigit-Elliott, S., and Züger, M.: Using Psychophysiological Measures to Assess Task Difficulty in Software Development. In *Proc. Int'l Conf. on Software Engineering (ICSE)*, pp.402–413 (2014).
 - [16] Ishida, T., Uwano, H., and Ikutani, Y.: Combining Biometric Data with Focused Document Types Classifies a Success of Program Comprehension. In the 28th

- IEEE/ACM Int'l Conf. on Program Comprehension (ICPC), pp.366–370 (2020).
- [17] Ishida, T., and Uwano, H.: Time Series Analysis of Programmer' s EEG for Debug State Classification. In the 5th Edition of the Programming Experience Workshop (PX), No.25, pp.1–7 (2019).
- [18] Klem, G. H., Luders, H., Jasper, H. H., and Elger, C.: The ten-twenty electrode system of the international federation, The Int'l Federation of Clinical Neurophysiology. *Electroencephalography and clinical neurophysiology*, Vol.52, pp.3–6 (1999).
- [19] 堀忠雄：生理心理学 - 人間の行動を生理指標で測る, p.265, 培風館 (2008).
- [20] Saby, J. N., and Marshall, P. J.: The utility of EEG band power analysis in the study of infancy and early childhood. *Developmental neuropsychology*, pp.253–273 (2012).
- [21] Lin, Y., Liu, Z., and Gao, X.: Alpha-band oscillation during speech recognition under different sensory conditions. In 2015 8th Int'l Conf. on Biomedical Engineering and Inf. (BMEI), pp.153–157 (2015).
- [22] Wróbel, A.: Beta activity: a carrier for visual attention. *Acta neurobiologiae experimentalis*, pp.247–260 (2000).
- [23] Malmivuo, P., Malmivuo, J., Plonsey, R., and Electroencephalography.: In *Bioelectromagnetism: Principles and Applications of Bioelectric and Biomagnetic Fields*, Chapter 13, pp.365–373 (1995).
- [24] Ahmadi, M., Michalka, S. W., Lenzone, S., Najafabadi, M. A., Bai, H., Sumich, A., Wuensche, B., and Billingham, M.: Cognitive Load Measurement with Physiological Sensors in Virtual Reality During Physical Activity. In Proc. the 29th ACM Symposium on Virtual Reality Software and Technology (VRST '23), No.29, pp.1–11 (2023).
- [25] Gevins, A., Smith, M. E., McEvoy, L., and Yu, D.: High-Resolution EEG Mapping of Cortical Activation Related to Working Memory: Effects of Task Difficulty, Type of Processing, and Practice. *Cerebral Cortex*, Vol.7, No.4, pp.374–385 (1997).
- [26] Klimesch, W.: EEG Alpha and Theta Oscillations Reflect Cognitive and Memory Performance: A Review and Analysis. *Brain Research Reviews*, Vol.29, No.2-3,

pp.169–195 (1999).

- [27] Klimesch, W.: Alpha-Band Oscillations, Attention, and Controlled Access to Stored Information. *Trends in Cognitive Sciences*, Vol.16, No.12, pp.606–617 (2012).
- [28] Niels, R., *Dynamic Time Warping: An Intuitive Way of Comparing Time Series Data*. Master's thesis, Radboud University Nijmegen (2004).