

ヒト脳内処理の解明に向けた取り組み

理学専攻・情報科学コース

2240647

川崎 春佳

1 はじめに

近年、ヒト脳の情報処理機構を解明する多くの研究が行われている。それらの研究では、様々なモデルを用いることによって様々な認知機能等に関する脳活動が調査されたり、ヒト脳活動の基礎となる脳活動パターンが調査されたりしている。このような背景を踏まえ、本研究では、動画視聴時の視覚・意味情報の局在と表現の調査と基礎となる脳活動パターンの調査を行う。

2 データ

使用したデータは情報通信研究機構脳情報通信融合センターにて取得された。脳活動データは、機能的磁気共鳴画像法 (fMRI) により取得された (3T MRI, TR 1 秒, ボクセルサイズ $2 \times 2 \times 2\text{mm}$)。fMRI 実験には被験者 7 名 (男 3 名, 女 4 名, 平均年齢 24.1 歳) が参加した。各被験者は 2 時間 40 分の音声付き自然動画を、固視 (スクリーン中央の固視点を注視する) 条件と自由視 (自由に視線を動かす) 条件で視聴した。各条件で得られた 2 時間 40 分のデータのうち、2 時間は訓練データとして用いた。残り 40 分は映像に 4 回の繰り返しがあり、それらを平均して 10 分にしたものを評価データとして用いた。また、1 秒ごとの各動画シーンに対して 5~6 名のアノテータから文章によるシーン記述も取得された。アノテータは日本語を母語とし、fMRI 実験には参加していない。

3 視覚・意味情報の局在と表現の調査

畳み込みニューラルネットワーク (CNN) を用いた研究では脳内での視覚情報の階層的な処理が確認されており [1, 2], 言語モデルを用いた研究では脳内での意味情報表現をモデル化可能であることが確認されている [3] が, その情報局在と情報表現には差異がある可能性が示唆された [4]。本研究では, 視覚・意味情報の大脳皮質上の情報局在と情報表現を視覚・言語・マルチモーダル情報を扱う様々なモデルを用いて調査する。概要を図 1 に示す。視覚・意味情報に関する脳活動をそれぞれ予測した後, 予測精度と予測脳活動を用いて Representational similarity analysis (RSA) を行う。

3.1 脳活動予測

被験者の体験内容に関する特徴量から線形回帰を用いて計測脳活動パターンと予測脳活動パターンが近づくように重みを学習させ, 体験内容から脳活動を予測する符号化モデルを作成する。

視覚情報に対応する画像特徴量は VGG16 の中間層 8 層 (MaxPool 層および Linear 層) と CLIP から抽出する。意味情報に対応する言語特徴量は CLIP, BERT_{BASE} の中間層 12 層, BERT_{LARGE} の中間層 24 層, GPT-2_{MEDIUM} の中間層 24 層から抽出する。モデルは全て事前学習済みのものを使用する。

抽出した特徴量の時系列を説明変数として脳活動の時系列を予測する符号化モデルをリッジ回帰により学習する。固視および自由視条件で各モデル各階層の特徴量からそれぞれ回帰し, 合計 140 個の符号化モデルを

作成する。学習した符号化モデルを用いて, 評価データでの計測脳活動と予測脳活動でピアソンの積率相関係数を求め, 各ボクセルの予測精度を評価する。False discovery rate で補正済みの p 値が有意 ($p < 0.05$) なボクセルを棄却する。

3.2 予測可能な脳領域と情報表現内容の分析

符号化モデルの予測精度と予測内容に対する RSA を行うことで, 大脳皮質上の視覚・意味情報の局在と表現内容の階層的特性を調査する。Representational dissimilarity matrix (RDM) 作成には相関距離を使用する。全ての符号化モデルで棄却され, かつ相関係数が 0 より大きいボクセルのデータを用いる。

(1) 予測可能な脳領域

全符号化モデルの予測精度を用いて (符号化モデル) \times (符号化モデル) の RDM を作成し, 全被験者で平均する。RDM を多次元尺度構成法で (符号化モデル) \times 3 に次元削減し, 3 次元空間上で可視化することで, 近くに配置される符号化モデル同士は予測可能な脳領域が類似しており, 離れた符号化モデル同士はそうではないということが見て取れるようにする。可視化結果を図 2 に示す。

(2) 空間的な情報表現内容

各符号化モデルの予測脳活動を用いて, (ボクセル) \times (ボクセル) の RDM を作成し空間的な情報表現を求める。次に, 全符号化モデルの RDM を用いて (符号化モデル) \times (符号化モデル) の RDM を作成し, 全被験者で平均する。その後, (1) と同様にして次元削減, 可視化することで, 符号化モデル間の空間的な情報表現の類似関係を見て取れるようにする。可視化結果を図 3 に示す。

(3) 時間的な情報表現内容

各符号化モデルの予測脳活動を用いて, (時間) \times (時間) の RDM を作成し時間的な情報表現を求める。その後は (2) と同様に分析を行う。可視化結果を図 4 に示す。

(4) 空間・時間両方を考慮した情報表現内容

各符号化モデルの予測脳活動を平坦化することで, 予測脳活動を (符号化モデル) \times ((ボクセル) \times (時間)) の形の行列に変形する。その後は (2), (3) と同様に分析を行う。可視化結果を図 5 に示す。

3.3 考察

実験結果より, 予測可能な脳領域の類似性は空間的な情報表現の類似性と類似しているが, 時間的な情報表現の類似性とは類似していないことが分かった。また, 情報表現の観点から深層学習モデルを比較すると, CLIP の画像特徴量は視覚情報表現と意味情報表現の中間的な脳内表現をモデル化できている可能性があることや, 言語モデル間で時間的な情報表現が異なる可能性があることが分かった。さらに, 視聴条件間で空間的な構造は異なるが, 時間的な構造は保持される可能性があることも分かった。

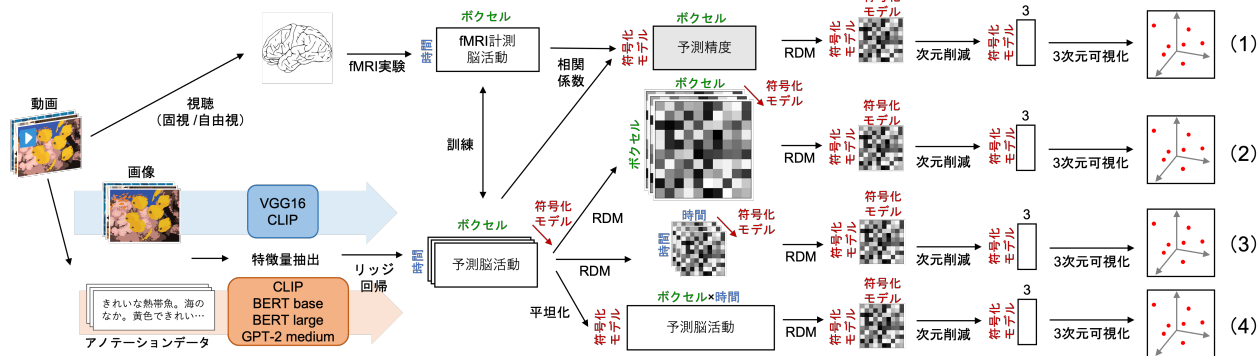


図 1: 視覚・意味情報の局在と表現の調査概要 (項番は 3.2 のものと対応する)

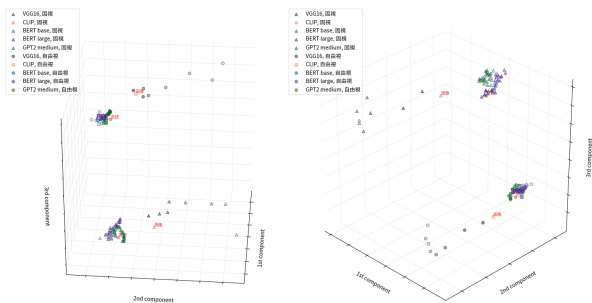


図 2: (1) 可視化結果

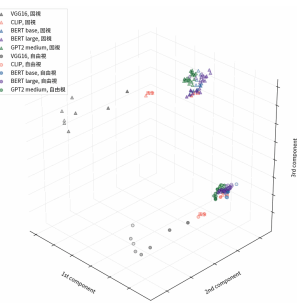


図 3: (2) 可視化結果

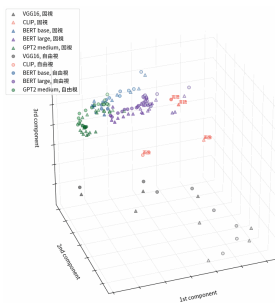


図 4: (3) 可視化結果

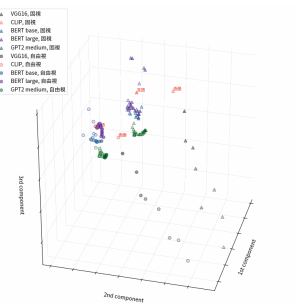


図 5: (4) 可視化結果



図 6: 基礎となる脳活動パターンの調査概要

表 1: 抽出した基底の解釈結果

順位	解釈
1	運動関係
2	安静時
3	音声言語処理等の聴覚
4	視覚と運動
5	痛み等大脳皮質以外に関連する処理
6	感情等の高次処理
7	形状認識に関する視覚
8	物体の運動認識に関する視覚
9	価値判断等の高次処理
10	ピッチ等の聴覚

4 基礎となる脳活動パターンの調査

ヒト脳活動の基礎となる脳活動パターンを調査する研究が行われている [5]. 本研究では、辞書学習を用いて基礎となる脳活動パターンの調査を行う。概要を図 6 に示す。脳活動データから辞書学習を行った後、代表的な基底がどのような脳活動であるか解釈する。

4.1 辞書学習

以下の実験では、自由視条件の脳活動データを使用する。訓練データとして 6 名の被験者の訓練データから 30000 秒分をランダムに抽出したものを、評価データとして他の 1 名の評価データを用いた。まず、大脳皮質を HCP-MMP1.0 により分割した脳領域ごとに、領域内全ボクセルのデータを平均する。次に、平均した訓練用脳活動データを用いて基底数 1000 で辞書学習を行う。そして、領域内平均した評価用脳活動データと学習した辞書を用いてスパースコーディングにより係数を求める。重要な基底を求めるために、評価データで求めた係数に対し、それぞれの基底に対応する全ての係数の L2 ノルムを求め、L2 ノルムが大きいものから 10 個に対応する基底を抽出する。抽出した基底をメタアナリシスツールの Neurosynth のデコーダに入力し、出力された単語のうち相関係数が高いものを用いて基底がどのような脳活動であるか解釈する。

4.2 結果と考察

抽出した 10 個の基底をそれぞれ Neurosynth に入力し、得られた結果を解釈することでどのような脳活動であるかを推測した結果を表 1 に示す。この結果から、それぞれの基底の解釈が少しずつ異なっており、辞書学習で基礎となる脳活動パターンを求められる可能性があることが分かった。また、抽出した基底は視覚や聴覚に関する脳活動が多かった。これは使用した脳活動が動画視聴時のものであるからではないかと考えた。

5 おわりに

本研究では、動画視聴時の脳活動について大脳皮質上の視覚・意味情報の局在と表現内容の特性を調査し、局在と様々な情報表現との関係性やモデル間の関係性、視聴条件間での差を確認した。また、基礎となる脳活動パターンも調査し、辞書学習によりそれを求めることができる可能性があることを確認した。動画視聴時の脳活動のみを扱った実験であるため、他の認知機能を含む脳活動に変更して実験を行いたい。また、分析に人間による解釈が入っているため、より定量的に判断できる分析手法に変更していきたい。

参考文献

- [1] Daniel L. K. Yamins, et al. Performance-optimized hierarchical models predict neural responses in higher visual cortex. *Proc Natl Acad Sci*, 111(23), pp.8619-24, 2014.
- [2] M. Eickensberg, et al. Seeing it all: Convolutional network layers map the function of the human visual system. *NeuroImage*, 152, pp.184-194, 2017.
- [3] S. Nishida, et al. Behavioral correlates of cortical semantic representations modeled by word vectors. *PLoS Comput Biol*, 17(6), e1009138, 2021.
- [4] H. Kawasaki, et al. Hierarchical processing of visual and language information in the brain. *Findings of AACL-IJCNLP 2022*, pp.405-410, Online only.
- [5] X. Chen, et al. Leading basic modes of spontaneous activity drive individual functional connectivity organization in the resting human brain. *Commun Biol*, 6(892), 2023.