

聴覚と視覚からなる言語刺激下での脳活動状態推定による言語処理領域の比較

理学専攻・情報科学コース

2040645

隅田 莉香子

1 はじめに

近年、脳波 (EEG) や脳磁図 (MEG)、機能的磁気共鳴画像法 (fMRI) など脳内状態の計測手法の発展に伴い、脳神経活動を観測した定量情報に基づき理解する研究が盛んに進められている。

また、Yamins ら [1] によって深層学習モデル階層的処理と脳内における視覚情報処理の階層的処理の間に相同性があることが示された以後、深層学習モデルを作業モデルとして援用することによりヒト脳内の情報処理機構の解明を目指す研究が盛んになっている本研究では言語からなる刺激 (以下、言語刺激) を与えられた際のヒト脳内の活動状態を調査する。

2 言語刺激による脳内状態推定

本研究では、音声発話が聴覚を通じて与えられた場合、および、同じ発話内容の書き起こし文章が視覚を通じて与えられた 2 つの場合について調査する。

2.1 実験設定

以下に示す 2 つの課題を検証する。

1. 傾聴と黙読を刺激とした際の脳活動状態の推定
2. 傾聴時と黙読時での脳内における情報処理領域の比較

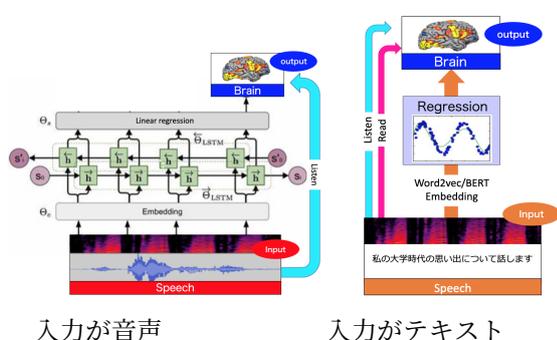


図 1: 本研究で用いたモデル概要図

言語モデル

本研究では、word2vec [2], BERT [3], RoBERTa [4] の言語モデルから得られた埋め込みベクトルを回帰モデルの入力に用いた。

脳活動データ

日本語話し言葉コーパス (CSJ)¹による会話音声傾聴時と同様の会話音声の文字起こし黙読時の被験者の血中酸素濃度に依存する信号 (BOLD 信号) を fMRI を用いて記録した脳神経活動データ $96 \times 96 \times 72$ ボクセルのうち大脳皮質に相当する 65,255 次元のデータを使用した。1 つあたりのデータセットは 12 分ほ

どで、train データと test データとで計 14 セット使用した。

脳活動状態推定モデル (入力がテキストの場合)

fMRI の観測遅延を考慮し 3~6 秒遅れとなる時点の 4 秒分を説明変数とした。回帰モデルに Ridge 回帰を採用した。全データの 9 割を訓練データ、1 割を評価データとし、訓練データに対して 5 分割交差検証を行い各ボクセルにおいて推定値のマクロ平均が一番高い正則化項 α の値を採用しモデルを構築した。これにより、ボクセルひとつずつに対して Ridge 回帰モデルを設定するとした。正則化項 α には次の 12 項を用いた。 $1.0 \times 10^2 \sim 10^7$, 2.5×10^4 , $5.0 \times 10^{-1} \sim 10^1$ 。また、評価指標としてピアソンの積率相関係数を用いて推定結果と評価データの相関関係を求め、帰無仮説「推定値と評価値の間には相関関係はない」の下に両側検定を $p = 0.05$ で行い、仮説が棄却されたボクセル群に対する FDR (False Discovery Rate) による補正を行なったものを対象に評価を行なった。最後に、UC Berkeley の Gallant Lab.²において開発された脳の可視化ツールである PyCortex³を使って相関係数の可視化を行った (図 3,4)。

脳活動状態推定モデル (入力が音声スペクトログラムの場合)

音声特徴量をスペクトログラムに変換するには、Python の音声処理ライブラリである LibROSA⁴を用いた。音声処理のためのパラメータ設定を表 1 に示す。回帰モデルとして、音声をスペクトログラムの系列と考え、テキストと同じように Bi-LSTM/Bi-GRU/Bi-RNN により双方向からの畳み込みからなる情報として特徴量を取得するモデルとなる。

表 1: 音声処理パラメータ設定

サンプリングレート	: 16000
メル周波数分解能	: 128
ホップレングス	: 1600
FFT 窓	: 4096

関心領域 (Region of Interest)

脳の関心領域 (ROI: Region of Interest) は、Destrieux の定義 [5] に従う。大脳皮質を機能的に 148 に分割しており、本研究において関心領域 (ROI) の 3 桁の ID 番号の内、最初の「2」は左脳を表し、「3」は右脳を表す。次の 2 桁は 1~74 までの関心領域を示している。

²<https://gallantlab.org/news/>

³<https://github.com/gallantlab/pycortex>

⁴<https://librosa.org/doc/latest/index.html>

¹<https://ccd.ninjal.ac.jp/csj/>

2.2 実験結果と考察

実験結果を以下のように示す.

1. 推定された脳内状態の相関係数の可視化
2. 可視化によって大局的に捉えた傾向を定量的に見る為に、それぞれの ROI 毎の相関性を以下の式によって算出した. そのうち上位 10 の ROI を抽出した.

各 ROI の相関性

$$= \frac{\text{ROI 内対象ボクセルの相関係数の総和}}{\text{当該 ROI の総ボクセル数}}$$

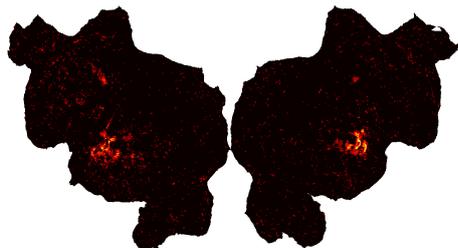


図 2: 音声スペクトログラムでの相関係数の可視化 (傾聴時-BiLSTM)

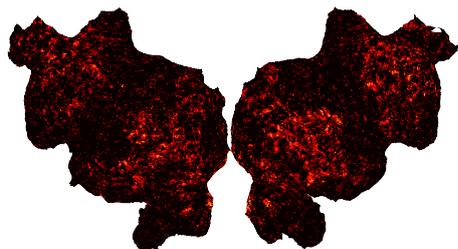


図 3: RoBERTa モデルでの相関係数の可視化 (黙読時)

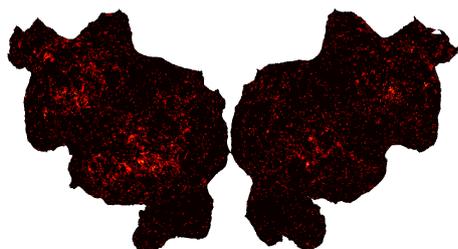


図 4: RoBERTa モデルでの相関係数の可視化 (傾聴時)

まず、脳活動計測時に与えられた刺激と同じものを入力とした回帰モデルの結果を図 2 と図 3 に示した. 結果として傾聴時には、音声処理を司る耳に近い領域に反応が見られ、黙読時には、単語理解を司る領域に強い反応が確認できた.

次に、傾聴時と黙読時でのモダリティ比較をするために、図 3 での同じテキスト入力を使用して、傾聴時における脳活動状態への回帰を行なった. この結果を図 4 に示す. また、傾聴時と黙読時においても全体的に word2vec の特徴表現より RoBERTa と BERT の方が高い相関と言語処理を司る脳領域に反応が確認された.

最後に、「傾聴と黙読のモダリティに依存しない言語処理領域」を特定するために、図 3 と図 4 の結果で同

じボクセルに反応が見られたものを抽出した. そして、2.2 節の冒頭に記述した「各 ROI の相関性」の算出方法に沿って、「傾聴と黙読のモダリティに依存しない言語処理領域」における上位 10 の ROI を表 2 に示した. 結果として、上位 10 のうち 7 つが言語を司る領域に反応が見られた.

表 2: 「傾聴と黙読のモダリティに依存しない言語処理領域」の ROI 上位 10

RoBERTa		
ID	名称	言語機能
273	上側頭溝	
234	上側頭回	音韻
212	弁蓋部	文法
352	下前頭溝の下部	
225	角回	単語
255	一次中間溝	単語
373	上側頭溝	
355	一次中間溝	単語
268	中心前溝の下部	
226	縁上回	単語

3 おわりに

本研究においては、傾聴と黙読により言語刺激が与えられた際の脳内状態を推定し、それぞれの特徴について考察を行なった. その結果として、傾聴時と黙読時ともに word2vec の特徴表現より RoBERTa や BERT の方が高い相関に加えて、言語処理を司る脳領域に反応が見られることがわかった. さらに、傾聴と黙読のモダリティに依存しない脳内意味理解領域は、ほとんどの反応領域が言語を司る部位に確認できた.

今後の課題としては、被験者を増やした実験の実施をしていきたい.

参考文献

- [1] Daniel L. K. Yamins, Ha Hong, Charles F. Cadieu, Ethan A. Solomon, Darren Seibert, and James J. DiCarlo. Performance-optimized hierarchical models predict neural responses in higher visual cortex. *Proceedings of the National Academy of Sciences*, Vol. 111, No. 23, pp. 8619–8624, 2014.
- [2] Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. Distributed representations of words and phrases and their compositionality. In C. J. C. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K. Q. Weinberger, editors, *Advances in Neural Information Processing Systems*, Vol. 26. Curran Associates, Inc., 2013.
- [3] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding, 2019.
- [4] Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. Roberta: A robustly optimized bert pretraining approach, 2019. cite arxiv:1907.11692.
- [5] Christophe Destrieux, Bruce Fischl, Anders M. Dale, and Eric Halgren. Automatic parcellation of human cortical gyri and sulci using standard anatomical nomenclature. *NeuroImage*, Vol. 53, No. 1, pp. 1–15, 2010.