

TSUBAME 共同利用 令和6年度 産業利用 成果報告書

利用課題名 EEG を用いた発話解読 AI の開発
英文: Development of EEG to Speech BCI Decoder
利用課題責任者 金井良太
First name Surname Ryota Kanai

所属 株式会社アラヤ
Affiliation Araya Inc.
URL www.araya.org

邦文抄録(300 字程度)

我々は、EEG を用いた発話解読 AI の実用化に向けた二つの重要な進展を報告する。1 つ目は、269 時間の EEG・音声のペアデータで学習した EEG エンコーダに対しゼロショット音声分類を行い、61.3%の精度を達成。データ量が精度に大きく影響することを示した。2 つ目は、被験者に課したタスクや電極配置が異なる EEG/EMG データをマルチタスク学習で処理するニューラルネットを提案し、EEG/EMG から silent speech で発話した単語を予測した際の正答率が健常者で 95.3%、神経変性症により発話困難な患者で 54.5%の精度を実現。特に健常者のデータから患者への転移学習により、単一被験者学習(健常者 70.1%、患者 13.2%)を大きく上回る性能を達成した。どちらも発話障害者支援の実用化に向けた有望な成果である。

英文抄録(100 words 程度)

We report two key advances toward practical EEG to speech BCI. First, zero-shot speech classification with 269 hours of EEG data achieved 61.3% accuracy, demonstrating a strong correlation between data volume and performance. Second, we propose a neural network that processes EEG/EMG data collected under different tasks and electrode configurations via multi-task learning, achieving 95.3% accuracy in healthy subjects and 54.5% in a speech-impaired patient with a neurodegenerative disorder. Notably, transfer learning from healthy participants to the patient significantly outperformed single-subject training (70.1% and 13.2%). These findings represent promising steps toward real-world BCI systems to support individuals with neurodegeneration-related speech impairments.

Keywords: brain-machine interface, transfer learning, EEG, speech recognition

背景と目的

ALS 等の神経変性症や、脳卒中後遺症により発声が困難な患者のコミュニケーションを補助する技術として、脳波(EEG)や筋電(EMG)を用いた発話解読のブレイン・コンピュータ・インターフェース(BCI)技術が注目されている。従来の侵襲的手法(EECoG, マイクロ電極)による発話解読 BCI は、高精度ながらも、外科的な手術を伴うため、感染症のリスクや心理的負担が大きい。そのため、非侵襲な代替手段として EEG/EMG による高精度な発話解読手法が求められる。

近年、言語モデルを中心に、データの量と AI モデルの精度に相関(スケーリング則)があることが知られているが、生体信号でノイズが大きいかつ、個人差や個人内でも日毎の差がある EEG においても、同様のスケーリング則が成立するかは不明であった。また、EEG は

取得コストが高く、患者の訓練用のデータを可能な限り少なくすることが実用的な BCI の開発には不可欠である。これら二点を検証するために、我々は、記録した大規模な発話中の EEG データを用いて、精度とデータ量間のスケーリング則の検証、並びに、多様な電極配置を扱うモデルを用意して、複数種の EEG/EMG デバイスで記録された大規模な健常者データから、神経変性症により発話が困難な患者データへの転移学習とその評価を行なった。

概要

1. 発話解読 BCI のスケーリング則評価概要

EEG を用いた発話解読 BCI のスケーリング則を評価するため、学習に使用するデータの総量を変化させながら以下のように学習と評価を行なった。発話中の EEG と音声のペアデータを 5 秒間のセグメントに分割

し、各セグメントの EEG の特徴量とペアとなる音声の特徴量(wav2vec 2.0 を使用)を類似させるように EEG エンコーダ(Conformer モデル)の対照学習(CLIP)を行なった(図 1)。モデルの性能指標として、テストデータに対するゼロショット分類性能(バッチサイズ=512)を用いた。本実験における実験参加者は健常者 1 名で、EEG 取得には 2 種類の脳波計測デバイス(g.PANGOLIN, g.SCARABEO)を用い、それぞれ独立にスケーリング則の評価を行った。

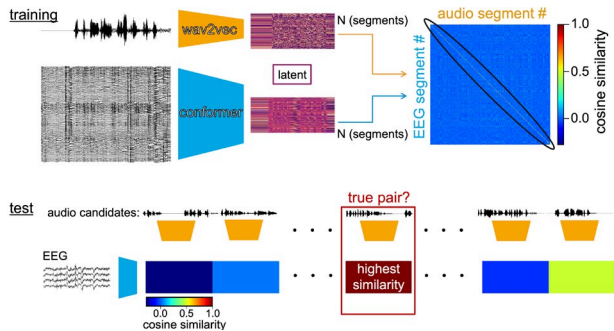


図 1 EEG から音声特徴量を予測するモデルの学習・評価のフレームワーク

2. 健常者から患者への転移学習評価概要

EEG/EMG を用いた発話解読 BCI の健常者から患者への転移学習可能性を検証するために、単語発話中の EEG/EMG から有声/無声発話した単語の分類(総単語数 64)の学習を以下の 3 条件で実施した。モデルは被験者毎の線形変換層を持つ Conformer モデルを用いた。本実験における健常者の実験参加者数は 8 名、患者の実験参加者数は 1 名であり EEG 計測デバイスには eego sports を用いた。

1. (被験者内): 各被験者個人のデータでのみ学習(各健常者 1.9 時間, 患者 1 時間)。
2. 全被験者のデータを混合したデータセットで学習(合計 16.1 時間)。
3. 本実験で収集した全被験者のデータに加え、さらに、異なるタスク・異なる EEG 計測デバイスで記録された発話中 EEG データを追加したデータセットで学習(合計 220 時間)。

評価指標には、テストデータを用いて、無声発話中の 64 単語の分類の精度を用い、各条件で精度を比較した。

結果および考察

1. 発話解読 BCI のスケーリング則評価結果

図 2 に示すように学習データを 1 時間程度から 100 時間に増やしていきそれぞれのデータ量でモデルの学習を行うと、データ量の増加とともにゼロショット分類精度が向上し、top-1 accuracy で 61.3% (512 フレーズ中) に達した。これは、リスニングタスク中の EEG と音声特徴量のゼロショット分類精度の SoTA (5% 程度)[1] を大幅に超えるものであった。使用した 2 つのデバイスの両方ともデータ量の増加に伴い同程度の精度向上が確認された。この成果は BCI Award 2024 にノミネートされた[2]。

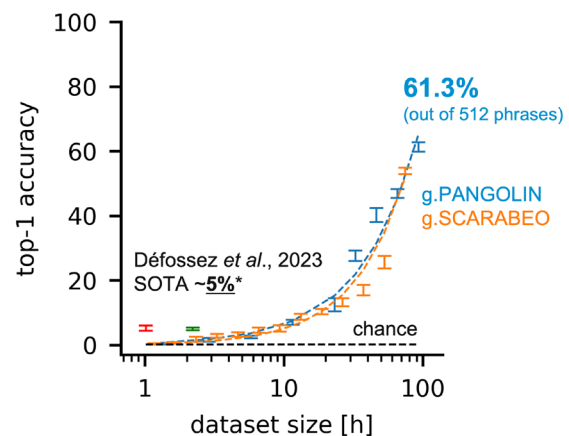


図 2 学習に用いた EEG・音声データサイズと zero-shot 分類性能の関係

2. 健常者から患者への転移学習評価結果

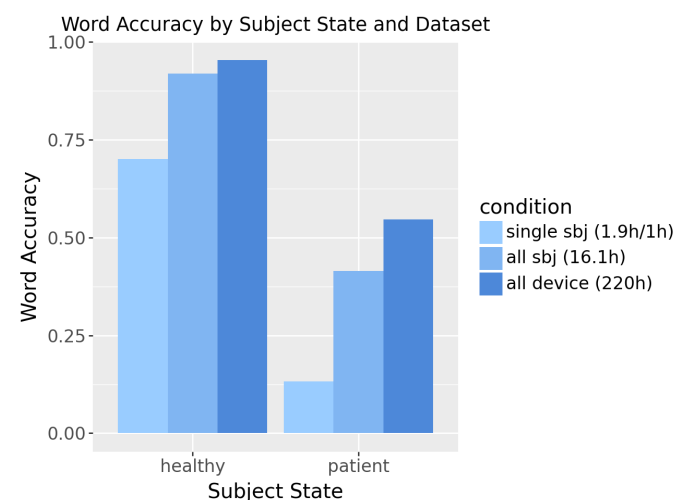


図 3 健常者・患者それぞれの、学習に用いたデータセットと単語分類精度

データセットを被験者内データから全被験者データ、異なるデバイスで記録された EEG を含むデータへと拡大していったときの、無声発話中の EEG/EMG から発話した単語分類タスクにおける健常者平均 (N=8) と患者

(N=1)の分類精度を図 3 に示す。健常者、患者ともに、被験者内データのみを使う場合と比較して、全被験者、デバイス混合のデータセットで学習した場合の精度が高い。また、データセットの大きさによる精度向上の効果(13.2%から 54.5%)は、患者データにおいて顕著であった。以上のことは、異種デバイス間での EEG/EMG データの転移学習が可能であることと、健常者データから、健常者とは発声方法や電極配置が異なる患者データにおいても転移学習が可能であることを示すものである。

まとめ、今後の課題

我々は、大規模な EEG と音声データを用いて、EEG を用いた発話解読 BCI のスケーリング則を示し、また、多様な記録デバイス・タスクで大規模に収集した健常者の EEG/EMG データから、患者の EEG/EMG データへの転移学習が可能なことを示した。このことは、BCI のエンドユーザーのキャリブレーションに必要なデータ量を最小限に抑えた、実用的な BCI システムの開発の可能性を示唆するものである。

今後は、より多様で大規模な EEG データを用いた学習により、よりロバストな BCI の開発を目指す。

参考文献

- [1] Défossez, A., Caucheteux, C., Rapin, J. *et al.* Decoding speech perception from non-invasive brain recordings. *Nat Mach Intell* **5**, 1097–1107 (2023).
- [2] Sato, A., Tomeoka, K., Horiguchi, I., Arulkumaran, K., Kanai, R., and Sasai, S., Large-scale data practicalizes EEG-based speech decoding. *BCI Awards* (2024)