

脳磁図を用いた音声提示時の脳活動解析について

山下 正人[†] 中沢 実[†]金沢工業大学[†]

1. はじめに

近年では、脳信号の計測や脳への刺激を与えることにより、脳と機械を直接つなぐ技術の Brain Machine Interface (BMI) に関連する研究開発が行われてきている。出力型 BMI は、脳内からの信号を外部の出力装置に送る技術である。この技術は、身体を思うように動かすことが出来ない難病患者の意思疎通や身体行動を手助けするために研究開発されている。難病患者の意思疎通は介護現場において重要な課題となっている。

この研究開発に対して、脳の神経活動を音声に変換する技術に注目されている。過去の研究では、純音や単音節を対象にした脳活動の解析が行われているが、連続した読み上げ音声を対象にした脳活動に対して、あまり研究は行われていない。本研究では、連続した読み上げ音声を対象にして、読み上げ音声を聴覚時の脳活動を非侵襲型の脳波計測手法である脳磁図 (MEG) を用いて行い、測定された脳活動から単音節ごと分類分けの可能性を検証する。

2. 関連技術

聴覚刺激における非侵襲型で測定された脳活動の反応に対する関連研究は既にいくつかの研究がある。

Khalighinejad, B ら [1] の研究では、EEG を用いて、連続音声における音素カテゴリへの誘発された神経反応の特性特徴付けを行っている。この研究では、音素に対する応答が音素の種類によって、音素開始後 50ms や 400ms の異なるタイミングで発生していることを示している。また、Chananel Braiman ら [2] の研究では、健常者と脳損傷患者の話し言葉の自然音声包絡線 (NSE) に対する EEG 反応の潜時および振幅を調査している。その結果、fMRI ベースの精神的画像タスクを実行できる脳損傷患者は、健常者と比較した際に、NSE 潜時において、統計的な有意差はなかったことを示している。山本 [3] の研究では、MEG を用いて聴覚誘発磁気反応を解析し、母音 /a/, /i/, /u/, /e/, /o/ に対する反応波形が各母音によって異なっており、個人差が大きいことが

Analysis of brain activity during voice presentation using magnetoencephalography

[†] Masato Yamashita, Minoru Nakazawa
Kanazawa Institute of Technology

7-1, Ohgigaoka, Nonoichi, Ishikawa, 921-8501, Japan

明らかとなっている。また、2 母音 (/a/ と /i/) を 180ms 間隔で提示した際に母音ごとに独立した反応波形が認められた。

これらの先行研究により、非侵襲型の脳活動測定における聴覚刺激の分類は可能であり、有益であると考えられる。

3. 脳活動の測定

本章では、聴覚刺激を用いた脳活動の取得方法及び聴覚刺激について説明する。MEG を用いた脳活動の測定には、同軸型グラジメータを用いて脳活動を測定する。チャンネル数は 160ch であり、サンプリング周波数は 1kHz である。バンドパスフィルタを適用し、0.1Hz から 200Hz の周波数帯域を取得する。測定は、金沢工業大学天池キャンパスのシールドルーム内で測定を行った。聴覚刺激を提示した際の脳活動の測定には、学習データの取得を 6 回、テストデータの取得を 6 回ずつ、各被験者に対して行う。

本研究の聴覚刺激には、JUST corpus [4] の音声データセットを用いた。JUST corpus は無響室で収録された一人の日本語女性話者の音声サンプリング周波数 48kHz で収録されている。

音声の種類は全部で 9 種類である。JUST corpus の音声データを後述するデータセットのサンプリング周波数と同一にするため、16kHz にダウンサンプリングしたものを利用した。学習データ用の聴覚刺激は、countersuffix26 (助数詞) 以外の 8 種類の中からランダムに選んだ音声をつなぎ合わせ、ランダムに並び替えた学習データ用の音声データ (240 秒間) を作成した。

テストデータ用の聴覚刺激は、助数詞の中から単位が「個、枚、本、冊、台、番」の 6 種類を単語ごとに分割し、ランダムに並び替えた音声データ (60 秒間) を作成した。音声のつなぎ合わせには、無声区間 (0.5 秒間) を挿入している。各試行で利用する音声データは共通であるが、各音声再生の順番はランダムである。

被験者として男子大学院生 1 名に協力してもらい、聴覚刺激提示時の脳活動測定を行なった。聴覚刺激の提示には、シールドルーム外から延長したエアチューブを用いて、被験者に提示を行なった。実験の開始前にエアチューブを耳につけた際に聴覚刺激を聞いてもらい、被験者が音声を聞き取りやすい音量に調整した。学習デ

ータとテストデータの取得は同一日に行なった。試行間に休憩を挟みながら実験を行なった。休憩間、被験者はシールドルーム外に出ることはなく、ベッドに横になっている状態のままである。試行の開始と終わりはシールドルーム内に取り付けたスピーカーを介し被験者に合図を送った。

MEGによって測定された脳活動を脳信号(brain wave)として定義する。目視で脳信号を確認し、ノイズが多く含まれているデータを解析の対象から除外した。

4. 解析手法・検証結果

脳信号を音素ごとに分類するために、連続音声に対して音素ごとの切り出しを行う。

音素の切り出しには、Julius音素セグメンテーションキット[5]を使用する。使用した音声コーパスの文章テキストと音声ファイルから、音素ごとの開始時間と終了時間を得る。本研究では、音素を母音(/a/, /i/ etc.)とCV音節(/ha/, /ki/ etc.)に焦点を当てるため、各母音と各CV音節の形で、脳信号を音素開始から-100ms~600ms区間で分割を行った。

分割の結果、多かった音節上位10種類(/i/, /u/, /N/, /shi/, /ru/, /no/, /ha/, /o/, /a/, /to/)の脳信号を解析対象とする。

脳信号の解析は、主に、前処理・特徴抽出・分類の3つに分けられる。

脳信号の前処理には、独立成分分析を用いた空間フィルタと帯域通過フィルタ、包絡線検出を適用する。空間フィルタにより、脳信号を独立成分に変換し、標準偏差が100を超えている独立成分をノイズと判断し、削除した。残った独立成分から再びチャンネル情報に再構成を行った後に、帯域通過フィルタ(70~180Hz)を適用する。帯域通過フィルタ後の各信号に対して、包絡線を抽出する。

脳信号の特徴抽出には、前処理後の各脳信号に対して標準偏差(SD)、二乗平均平方根(RMS)、総和(SUM)の3種類を各チャンネルから抽出する。全チャンネルに対して、特徴抽出を行うため、特徴量は3(種類)×160(ch)=480個となる。

分類器の作成には、SVM(Support Vector Machine)を用いる。本研究では、SVMを別々のカーネルを用いて6種類作成し、分類性能を比較する。SVMで利用するカーネルの種類は線形・二次・三次・ガウス(スケール:5.5, 22, 88)の6種類である。本研究では、10種類の分類を行う必要がある。しかし、SVMは本来2種類の分類に適した手法である。そのため、多クラス分類に対応させるため、マルチクラスの1対1クラス方

式を適用することによって、10種類のクラス分類を行う。

1つの音節から200個の脳信号をランダムに選び、SVMの訓練と検証を5交差検証によって行った。各SVMの精度は線形:9.5%, 2次:11.3%, 3次:10.7%, ガウス(スケール:5.5):9.2%, ガウス(スケール:22):10.7%, ガウス(スケール:22):10.2%であった。最も精度が高かったのは、2次SVMの11.3%であった。これは、チャンスレベルの10%をわずかに上回っている。

5. まとめ・今後の課題

本研究では、連続した読み上げ音声を対象にして、読み上げ音声を聴覚時の脳活動を非侵襲型の脳波計測手法である脳磁図(MEG)を用いて行い、測定された脳活動から単音節ごと分類分けの検証を行った。脳活動の測定を行い、脳活動を音素ごとに切り分け、前処理・特徴抽出を適用し、6種類のSVMを適用することによって、分類精度を検証した。検証の結果、2次SVMを用いた時に分類精度が11.3%とチャンスレベルをわずかに上回る結果を得た。

今後の課題として、分類精度の向上が挙げられる。そのため、前処理・特徴抽出手法の変更や、チャンネル情報の選択などが挙げられる。

文 献

- [1] Bahar Khalighinejad, Guilherme Cruzatto da Silva, and Nima Mesgarani. Dynamic encoding of acoustic features in neural responses to continuous speech. *Journal of Neuroscience*, Vol. 37, No. 8, pp. 2176- 2185, 2017.
- [2] Chanel Braiman, Esteban A. Fridman, Mary M. Conte, Henning U. Voss, Chagit S. Reichenbach, Tobias Reichenbach, and Nicholas D. Schiff. Cortical response to the natural speech envelope correlates with neuroimaging evidence of cognition in severe brain injury. *Current Biology*, Vol. 28, No. 23, pp. 3833 - 3839.e3, 2018.
- [3] 山本智矢. 脳磁図による言語知覚の研究. *音声言語医学*, Vol. 35, No. 3, pp. 289-294, 1994.
- [4] Ryosuke Sonobe, Shinnosuke Takamichi, and Hiroshi Saruwatari. JSUT corpus: free large-scale japanese speech corpus for end-to-end speech synthesis. *CoRR*, Vol. abs/1711.00354, 2017.
- [5] Julius 音素セグメンテーションキット <https://julius.osdn.jp/index.php?q=ouyoukit.html> (最終閲覧日: 2020年1月8日)