

音声想起分類のための電流源推定に基づく 脳磁図データ拡張の検討*

☆能勢幸樹, 矢野肇, 高島遼一, 滝口哲也 (神戸大), 中川誠司 (千葉大)

1 はじめに

ブレイン・コンピュータ・インターフェイス (brain computer interface: BCI) は発話や身振りが困難な身体障害者のための意思伝達手段として研究開発が盛んに行われている。近年では、自由度の高いコミュニケーションが可能な、頭の中で想起した音声を脳活動から読み取る BCI の研究が注目されている [1]。

我々はこれまで、音声想起時の脳磁図 (magnetoencephalography: MEG) を調査し、機械学習を用いて想起音声の識別を行ってきた [2, 3, 4, 5]。先行研究 [3, 4] では、複雑な識別関数を学習可能な畳み込みニューラルネットワークが用いられたが、パラメータ数は古典的な機械学習手法よりも圧倒的に多く、実用に耐えうる高い精度を持つ識別器を個人ごとの少ない MEG データを用いて安定して訓練することは容易ではない。

一般的に被験者が行うタスクごとに脳波の特徴は異なり、タスクごとに効果的なデータ拡張手法が異なることが示されている [7]。我々は前稿 [6] で、音声想起時の MEG データに対するデータ拡張を調査し、MEG の時間波形をあまり変化させない手法が想起音声の分類精度を向上させることを示した。

脳活動データに対するデータ拡張では時間、周波数、空間領域でのデータ拡張がよく用いられる。これらは頭外に設置されたセンサーで計測される信号に対する処理だが、それらの信号源は脳内に位置する。このため、脳内の信号に対して直接処理を行うことでより効果的なデータ拡張を行える可能性がある。そこで本研究では、音声想起時の脳活動データの分類精度を向上させるための電流源推定に基づくデータ拡張手法を調査した。音声想起時の MEG データに対し、電流源推定に基づいたデータ拡張手法を適用し、想起された単語音声の識別をする軽量のニューラルネットワークモデルを被験者ごとに訓練した。データ拡張を用いずに訓練したネットワークとデータ拡張を用いて訓練したネットワークの分類精度を比較して、音声想起時の MEG の分類に有効なデータ拡張手法を検討した。

2 MEG データセット

先行研究 [2] で収録された音声想起時の MEG データを用いた。このデータは 8 名の被験者 (男性 7 名, 女性 1 名, 20–40 歳) の MEG データからなる。被験者は、3 種類の日本語単語 (“あまぐも”, “いべんと”, “うらない”) のうちの 1 つを 2 回聴取した後、音声を聞いた通りに想起する試行を繰り返す。音声の持続時間は 800 ms で、呈示間隔は 500 ms であった。この試行は単語ごとに少なくとも 100 回行われた。試行中の MEG は被験者の頭を覆うように配置された 122 個のセンサで計測された。

計測された MEG の前処理として、サンプリング周波数が 200 Hz となるようにダウンサンプリングし、1 Hz 以下の低周波成分、1000 fT/cm を超える変動、及び眼球運動に由来するアーティファクトを除去した。頻繁に異常な信号が観測された MEG センサの信号は取り除き、他の正常なセンサの信号を元に補間した。1 回目の音声呈示及び想起のタイミングを基準に -100 – 900 ms の信号を音声聴取時及び想起時の MEG として切り出した。最後に、MEG 波形の振幅を ± 100 fT/cm が ± 1 となるように正規化した。

3 電流源推定の定式化

3.1 MEG の順問題

本稿では、脳内の領域が M 個の点にメッシュ化され、電流源はメッシュ点ごとに配置された微小電流 $\mathbf{j}_i \in \mathbb{R}^3 (1 \leq i \leq M)$ で表されるとする。すべてのメッシュ点における微小電流を並べたベクトルを $\mathbf{j} = [\mathbf{j}_1^\top, \dots, \mathbf{j}_M^\top]^\top \in \mathbb{R}^{3M}$ と表す。電流分布 \mathbf{j} が与えられた時、頭外の N 個の MEG センサでの磁場 $\mathbf{b} \in \mathbb{R}^N$ は次の式で表される。

$$\mathbf{b} = \mathbf{L}\mathbf{j} \quad (1)$$

ここで、 \mathbf{L} はリードフィールド行列である。 \mathbf{L} はメッシュ点とセンサの位置、頭内の導電率から求めることができ本研究では既知とする。

3.2 MEG の逆問題

一般的に、脳内をメッシュ化した場合、電流源のパラメータ数はセンサ数に対して非常に大きくなる

*Data augmentation based on source estimation for imagined speech classification. by NOSE, Kōki, YANO, Hajime, TAKASHIMA, Ryōichi, TAKIGUCHI Tetsuya (Kobe Univ.) and NAKAGAWA, Seiji (Chiba Univ.).

ため、式 (1) の \mathbf{b} の観測値 \mathbf{b}_{obs} と \mathbf{L} が与えられたとしても、 \mathbf{j} を一意に決めることはできない。代表的な MNE に基づく逆問題の解法は、 \mathbf{j} に制約を与え、式 (1) の順問題と観測値の間の誤差を最小化することで \mathbf{j} を一意に求める。 \mathbf{j} の推定値 $\hat{\mathbf{j}}$ は以下で与えられる。

$$\hat{\mathbf{j}} = \underset{\mathbf{j}}{\operatorname{argmin}} \|\tilde{\mathbf{b}}_{\text{obs}} - \tilde{\mathbf{L}}\mathbf{j}\|_2^2 + \lambda \mathbf{j}^\top \mathbf{S}^{-1} \mathbf{j} \quad (2)$$

ここで、 $\tilde{\mathbf{b}}_{\text{obs}}$ はセンサにおける観測ノイズの共分散行列が単に行列となるように白色化された観測値、 $\tilde{\mathbf{L}}$ は白色化されたリードフィールド行列を表す。 \mathbf{S} は電流源のパラメータの共分散行列である。

sLORETA [8] や eLORETA [9] などの一部の手法は、ノイズのない理想的な条件下や、特定の構造を持ったノイズのある条件下でも、脳内の単一電流源の位置を正確に推定することができるが示されている [9]。

4 電流源推定に基づくデータ拡張

本研究では、音声想起時の MEG データに対し、電流源推定に基づくデータ拡張を適用した。まず、実際の MEG データに対し式 (2) で電流源推定を行った。そこで得られた電流源 \mathbf{j} に対し以下で述べるデータ拡張を適用し、適用後の電流源を式 (1) で MEG データに戻すことで拡張後のデータを得た。電流源領域で適用した手法は、Rommel らの研究 [7] と前稿 [6] を参考にした。各データ拡張手法について以下に述べる。

4.1 ガウスノイズ付加 (Gaussian noise) [11]

各時間サンプルが独立に平均 $\mathbf{0}$ 、分散 $\sigma^2 \mathbf{I}$ の正規分布に従うランダムベクトルを確率 p で全ての電流ベクトルに加算する。このデータ拡張によって脳信号に含まれるノイズに対して頑健なモデルの学習が期待できる。

4.2 Largest current scaling

すべての電流源の中で瞬時的な強度が最大となる電流源のみに正の定数 σ を乗算する。活発な電流源を強調することで、音声想起に関連する脳活動が強調されることが期待される。

$$l_{\max} = \operatorname{argmax}_l \max_t |\hat{\mathbf{j}}_l(t)| \quad (3)$$

$$\hat{\mathbf{j}}_{l_{\max}}^{\text{aug}}(t) = \sigma \cdot \hat{\mathbf{j}}_{l_{\max}}(t) \quad (4)$$

4.3 Source dropout

各電流源の時系列 $\hat{\mathbf{j}}_l(t)$ を確率 p で 0 にする。なお、信号を 0 にした電流源の数に応じて MEG 波形の振幅が小さくなるため、拡張前と拡張後の電流源の強

度の平均が一致するように定数をかける。この処理によって、分類器が特定の電流源に由来する信号成分へ過学習することを抑制することが期待される。

4.4 Threshold dropout

すべての電流源に共通の閾値を計算し、閾値以下の最大瞬時強度を持つ電流源を 0 にする。閾値は以下の式によって求める。

$$\text{threshold} = p \cdot \max_{l,t} |\hat{\mathbf{j}}_l(t)| \quad (5)$$

また、 p は $[0, 1]$ の区間で設定される。ノイズに成りうる比較的小きな信号を除去することによって、ノイズの少ないデータが得られると考える。

4.5 周波数シフト (frequency shift) [10]

全ての電流源 \mathbf{j}_l の各要素の時系列信号のスペクトルを一様分布 $\mathcal{U}(-\Delta f_{\max}, \Delta f_{\max})$ からサンプリングしたランダムな周波数 Δf だけシフトする。各電流源の各要素の時系列信号を $x(t)$ 、その解析信号を $x_a(t) = x(t) + i\mathcal{H}[x](t)$ とする。ここで $i = \sqrt{-1}$ で、 $\mathcal{H}[x](t)$ は $x(t)$ のヒルベルト変換である。 $x(t)$ を周波数シフトした信号 $x_{\text{freqshift}}(t)$ は次のように計算される。

$$x_{\text{freqshift}}(t) = \operatorname{Re} [x_a(t) \cdot e^{i2\pi\Delta f \cdot t}] \quad (6)$$

ここで、 $\operatorname{Re}[\cdot]$ は複素数の実部を取り出す演算である。

このデータ拡張は、一部の脳信号が持つ特定の周波数のピークが変動する場合に有効である。

4.6 Fourier transform surrogate [12]

各電流源の各要素の時系列信号のフーリエ位相スペクトルを周波数ごとに一様分布 $\mathcal{U}(0, \Delta\phi_{\max})$ からサンプリングしたランダムな位相 $\Delta\phi$ だけシフトする。各周波数における $\Delta\phi$ は全電流源で共通である。各電流源の各要素の時系列信号 $x(t)$ のフーリエ変換を $\mathcal{F}[x](f)$ とすると、位相をランダム化した信号 $x_{\text{fts surrogate}}(t)$ のフーリエ変換は次の式で表される。

$$\mathcal{F}[x_{\text{fts surrogate}}](f) = \mathcal{F}[x](f) e^{i\Delta\phi} \quad (7)$$

処理後の時間信号 $x_{\text{fts surrogate}}(t)$ は式 (7) の逆フーリエ変換によって得られる。

このデータ拡張によって元の信号と同じ振幅スペクトルを持つ異なる波形信号を作ることができるため、周波数成分を重視するように学習させることができると考えられる。

5 評価実験

5.1 実験条件

音声想起時の MEG の分類モデルとして、EEGNet [13] を被験者ごとに訓練した。EEGNet は一般的な畳

み込みの代わりに depthwise separable convolution を用いた軽量なネットワークであり、データ数が少量でも訓練が可能である。EEGNet は時間方向の畳み込み層、センサ方向の depthwise convolution 層、時間方向の separable convolution 層、及び全結合層からなる。本研究では、EEGNet の構造を変更し、separable convolution 層を 2 層にして用いた。

訓練したモデルを 10 分割交差検証によって評価した。評価データセット以外のデータセットの 80% を訓練データセットに、20% を検証データセットに分割した。データセットの分割時、各分割内の試行が時間的に連続するように分割した。学習されたモデルの評価指標には macro F1 score を用いた。この指標はクラスごとに算出した F1 score を平均したものである。EEGNet は最大で 200 エポック訓練し、検証データセットに対して macro F1 score が最大となるモデルを選択した。

電流源推定には sLORETA を用いた。すべてのデータ拡張手法は予め訓練データセットのランダムに選ばれた 20% のデータに適用され、適用前と適用後のデータを集めた新たな訓練データセットがモデルの訓練に用いられた。

5.2 結果

データ拡張適用前と各データ拡張適用後の右側頭付近の MEG チャネルの信号波形を Fig.1 に示す。音声想起時の MEG データにデータ拡張を適用せずに訓練したモデルの macro F1 score と、データ拡張を適用した時の macro F1 score の改善量を Table 1 に示す。MEG データに対して多くの被験者で macro F1 score が向上したデータ拡張手法は、Threshold dropout と Fourier transform surrogate であった。一方で、他のデータ拡張手法も、macro F1 score を向上させたものの、Threshold dropout や Fourier transform surrogate と比べると、向上幅が小さかった。また、被験者ごとの結果は、データ拡張手法を適用した際に被験者 5, 7, 8 で macro F1 score が低下または向上幅が小さくなることが多く、被験者 4, 6 は他の macro F1 score が向上している被験者と比較して向上幅が大きくなるが多かった。

5.3 考察

$p=0.05$ を用いた Threshold dropout を適用した際、全被験者で分類精度が向上した。しかし、他の各データ拡張手法の効果は被験者の間で一貫していなかった。被験者間での違いは音声想起時の脳活動の違いによるものと考えられる。特に被験者 5, 7, 8 は分類精度が低下することが多く、分類精度が向上しているデータ拡張手法であっても多くの場合、他の被験者に

比べて向上幅が小さかった。よって、この 3 人の被験者の MEG データにはモデルの学習において妨げとなる特異な信号やノイズが含まれていたと考えられる。

Threshold dropout は全被験者で分類精度が向上した。これは比較的小さい信号を持つ電流源を 0 にしたことにより、重要でない信号が除去され、音声の想起に関連する電流源から MEG データを再構成できたためと考えられる。

Fourier transfer surrogate は前稿の結果 [6] とは異なり、被験者 8 以外で分類精度が向上し、低下した被験者についても低下幅は小さかった。これは、 $\Delta\phi$ を小さな値に設定していたため、ランダムにシフトされる位相も小さいものになり、元のデータセットに含まれるような MEG データが生成されたためと考えられる。

Largest current scaling は被験者 5, 8 以外で分類精度が向上した。最も強度の大きい電流源を強調したことで、音声想起に関連した MEG の成分が強調され、モデルが想起音声の識別に適した特徴量を抽出できるように訓練された可能性がある。

Source dropout は Threshold を用いたものと異なり、MEG データを再構成する際に、音声の想起に関連する電流源も 0 にしてしまうため、Threshold dropout よりも分類精度が下がってしまったと考えられる。ただし、0 にされる電流源によっては、拡張後の MEG 波形の振幅が非常に小さくなることが考えられ、元の振幅の平均と同じになるように定数をかけたことで、多くの被験者で分類精度を低下させるという結果には至らなかったと考えられる。また、電流源推定によって得られる電流源は局所的に類似しているため、0 にされた電流源のまわりの電流源から、失われた情報を補うことが可能であったことも考えられる。

周波数シフトは、被験者 5, 7 以外で分類精度が向上した。前稿 [6] では、同様のデータ拡張手法を MEG データに直接適用しており、分類精度が多くの被験者で大きく低下していた。今回の結果は、周波数のシフト幅 Δf を小さく設定したことが大きな要因であると考えられる。

Gaussian noise は、被験者 8 で分類精度が大きく低下し、それ以外の被験者で分類精度が向上した。精度が向上した被験者に関しては、脳信号に含まれるノイズに対してより頑健なモデルを学習できたといえる。

電流源空間の周波数領域でデータ拡張を行う場合、電流源の波形に位相や周波数に小さい変化を加えることは、想起音声の分類に有効である可能性が示唆された。また、今回用いたデータ拡張手法は Fig.1 の

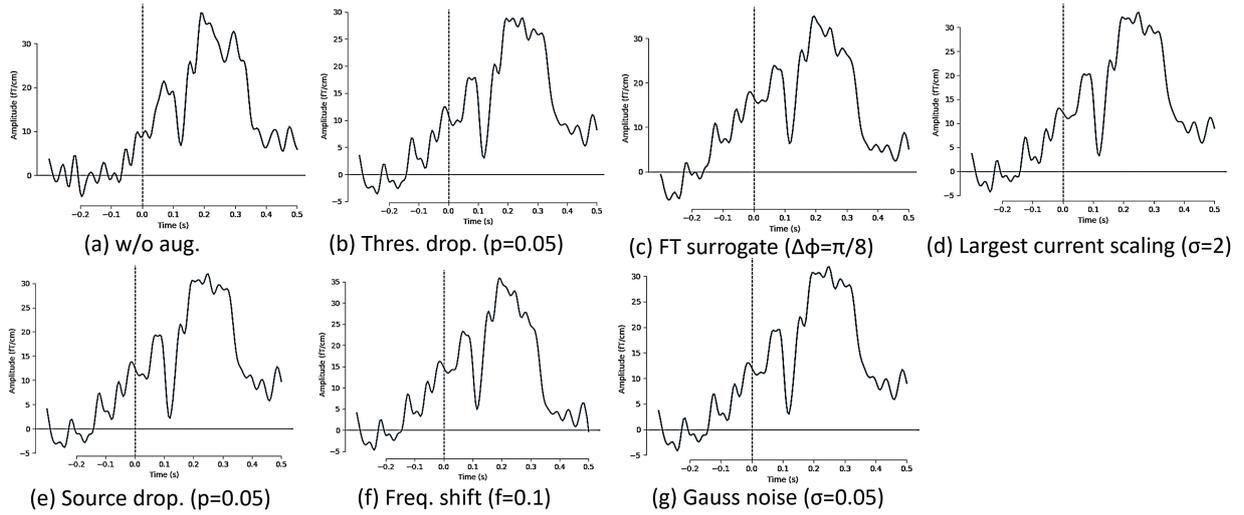


Fig. 1 (a) Observed and (b-g) reconstructed MEG waveforms at an MEG channel around the right temporal area.

Table 1 Amount of the macro F1 score improvement.

Subject	1	2	3	4	5	6	7	8	Mean
w/o augmentation	35.36	29.78	32.28	40.12	41.60	34.44	33.72	42.02	36.16
(1) Theres. drop. ($p = 0.05$)	+3.81	+4.18	+6.16	+2.02	+3.00	+8.96	+3.18	+2.03	+4.17
(2) FT surrogate ($\Delta\phi_{\max} = \pi/8$)	+6.36	+0.80	+6.43	+7.26	+1.65	+4.43	+0.55	-0.30	+3.40
(3) Largest current scaling ($\sigma = 2$)	+3.08	+6.19	+4.17	+6.82	-0.50	+6.13	-2.14	+0.16	+2.99
(4) Source drop. ($p = 0.5$)	+6.10	+1.98	+1.64	+6.25	+0.58	+7.15	+0.03	-0.16	+2.95
(5) Freq. shift ($f_{\max} = 0.1$)	+1.84	+3.89	+7.31	+5.71	-1.20	+4.11	-0.80	+1.32	+2.77
(6) Gauss noise ($\sigma = 0.05$)	+2.42	+3.09	+3.78	+5.64	+1.53	+3.56	+1.93	-1.94	+2.50

ように元の MEG 波形の概形をある程度保ちつつ異なる波形を生成することができたため、多くの被験者で精度が向上したと考えられる。

6 おわりに

本研究では、音声想起時の MEG が収録されたデータセットに対し、電流源推定に基づいたデータ拡張手法を適用し、想起音声の分類に有効なデータ拡張手法を調査した。その結果、電流源空間での Threshold dropout と Fourier transform surrogate を用いたデータ拡張が分類精度の向上に寄与した。電流源空間におけるデータ拡張は、音声想起に関連する電流源の信号を直接操作することで、より頑健な分類モデルを訓練できるデータを生成する可能性が示唆された。

謝辞 本研究の一部は、JSPS 科研費 JP22K18626 の支援を受けて実施された。

参考文献

- [1] C. H. Nguyen *et al.*, J. Neural Eng., 016002, 2018.
- [2] S. Uzawa *et al.*, IEEE EMBC, 2542–2545, 2017.

- [3] 矢野ら, 音講論 (春), 507–510, 2020.
- [4] 山名ら, 音講論 (春), 517–520, 2023.
- [5] 矢野ら, 音講論 (春), 645–648, 2024.
- [6] 矢野ら, 音講論 (秋), 915–918, 2024.
- [7] C. Rommel *et al.*, J. Neural Eng., 066020, 2022.
- [8] R. D. Pascual-Marqui *et al.*, Methods and Findings in Experimental and Clinical Pharmacology, 24(D):5–12, 2002.
- [9] R. D. Pascual-Marqui *et al.*, Phil. Trans. R. Soc. A, 369, 3768–3784, 2011.
- [10] C. Rommel *et al.*, ICLR, 2022.
- [11] F. Wang *et al.*, MultiMedia Modeling, 82–93, 2018.
- [12] J. T. C. Schwabedal, *et al.*, arXiv:1806.08675, 2019.
- [13] V. J. Lawhern *et al.*, J. Neural Eng., 15, 056013, 2018.