

## EEGNet を用いた脳磁界データによる想起時の音声識別の検討\*

☆山名莉央, 北条直樹, 矢野肇, 滝口哲也, 高島遼一(神戸大), 中川誠司(千葉大)

### 1 はじめに

身体障害者の生活支援の手段として, 音声認識や行動認識を用いた機械制御・意思伝達技術が用いられる。しかし, 発話や身振りでのコミュニケーションが困難な障害者は, 音声認識技術を有効に活用することができない。そこで, 脳活動を用いて機械制御をおこなうブレイン・コンピュータ・インターフェイス (BCI) の利用が期待される。

従来の BCI 開発では, 注意を向けた低頻度刺激に対してのみ出現する誘発反応 (P300) によって, 限定的な意思伝達を行う“P300 スペラー型 [1]”の開発例が多い。しかし, P300 スペラー型では予め用意された選択肢の中から使用者の意思が判別される仕組みとなっており, 自由度の高い意思伝達はできない。ユーザの意思をより汎用的に他者に伝達するためには, 脳内でユーザが思考したままに認識するようなシステムを開発することが望ましい。そこで我々は, ユーザの意思をより汎用的に認識する BCI の実現に向け, 脳磁界データから音声想起の識別を試みた。

脳磁界データにおける想起音声の識別精度向上が困難な理由として, 音声想起時の脳活動には音声聴取時に比べて明らかな音響特徴の鈍化・欠落が生じており, 単語を識別するのは容易ではないことや, 被験者実験を伴うデータ計測が長時間行えないことにより, データ数が極端に少なくなってしまうことがあげられる。

我々はこれまで, 脳磁界データにおける音声想起に伴う信号の微弱さ, データの少なさに対する特徴量の次元数の膨大さを克服するため, 識別的なチャンネル選択を行うことを目的として, SENets を組み込んだ CNN を実装した [3]。しかし, 実用上十分な識別精度を得られたとは言い難い。

本研究では, 脳磁界データにおける想起音声識別の精度を向上させるため, EEGNet [2] を用

いた手法を提案する。EEGNet は神経生理学的信号から解釈可能な抽象的特徴を抽出するだけでなく, 少ないデータセットで信頼性の高いモデルを構築することが可能である。そのため, 様々な BCI タスクにおいて優れた性能を示している手法である。EEGNet が他の BCI タスクと同じようにデータ数が少ない音声想起タスクに対しても有効であることが期待される。音声想起に対しての EEGNet の有効性を示すために, 音声想起を行なった実際の脳磁界データを用いて実験を行った。

### 2 脳磁界計測と実験内容

5名の聴覚健常者に対し, 3パターンの単語音声 (“あまぐも”, “いべんと”, “うらない”) を用いて実験を行った。3単語のうち1単語がランダムに選択され, 文字刺激として3回呈示を行った [4]。この時, 1回目と2回目には同時に音声刺激も呈示した。被験者には, 1回目, 2回目の単語音声を聴取した後, 3回目は音声を流さず文字呈示のみを行い, 3回目の文字呈示のタイミングで1,2回目で聴取した音声を想起するよう求めた。1試行の概要を Fig. 1 に示す。

音声刺激には, 親密度音声データベース (FW03, NTT-AT) に含まれる女性話者音源 (fto) を利用した。刺激呈示時間は 800ms であり, 解析対象の脳磁界データの解析時間は, 想起開始時の合図から -200ms から 800ms とした。

脳磁界計測に, 122ch 全頭型脳磁界計測システム (Neuromag-122<sup>TM</sup>:Neuromag, Ltd.) を用いた。計測した脳磁界データは 0.03-100Hz のアナログフィルタを適用した後, サンプリング周波数 400Hz で A/D 変換を行った。同時に, ノッチフィルタを適用した。

\*Sound classification of brain magnetic fields using EEGNet. by YAMANA, Rio, HOJO, Naoki, YANO, Hajime, TAKASHIMA, Ryoichi, TAKIGUCHI Tetsuya (Kobe Univ.) and NAKAGAWA, Seiji (Chiba Univ.).

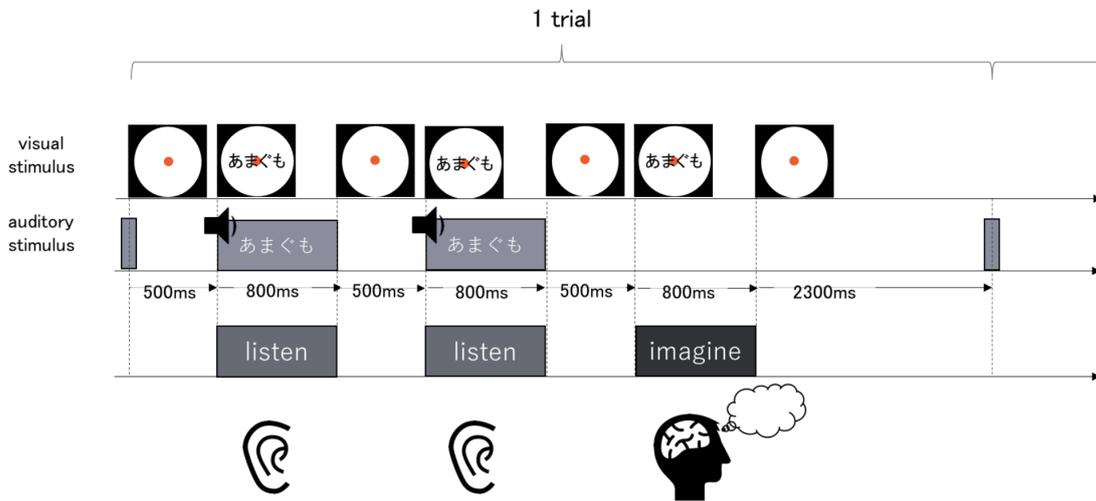


Fig. 1 Schematic diagram of the task.

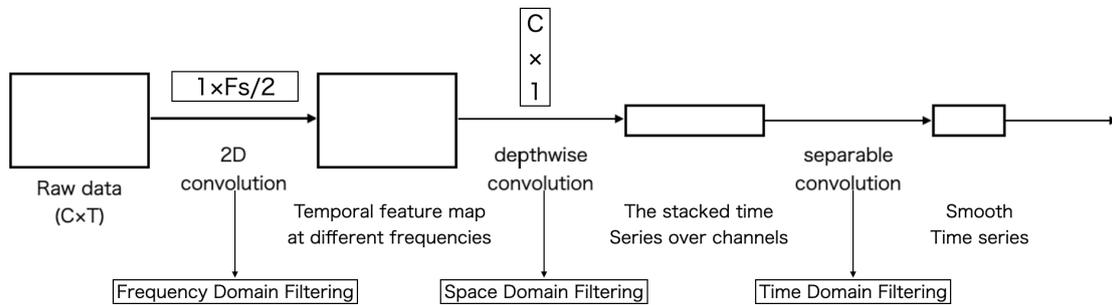


Fig. 2 Overall visualization of the EEGNet architecture.

### 3 EEGNet を用いた音声想起識別

#### 3.1 EEGNet

EEGNet は様々な BCI タスクに適用可能な畳み込みニューラルネットワークである。例えば、P300, 視覚誘発電位, エラー関連陰性反応 (ERN), 運動関連皮質電位 (MRCP), 感覚運動リズム (SMR) などで優れた性能を実証されている [2].

EEGNet は入力を Raw データとし, 時間方向の畳み込み, Depthwise 畳み込み, Separable 畳み込みからなるネットワークである。構造を Fig. 2 に示す [5]. まず時間方向の畳み込みで周波数フィルタを学習し, Depthwise 畳み込みで各特徴マップに個別に, 周波数に応じた空間フィルタを学習する。最後の Separable 畳み込みで, 各特徴マップを時間方向に個別に学習する畳み込みと, 特徴マップを最適に加算する畳み込みを組み合わせたものである。

つまり, Depthwise 畳み込みはチャンネル方向の計算を行わず, 空間方向のみの計算を行う畳み込みであり, 入力と出力のチャンネル数の変化はない。Separable 畳み込みは Depthwise 畳み込みと Pointwise 畳み込みから構成されている。Pointwise 畳み込みは  $1 \times 1$  畳み込みと言われ, 空間方向には  $1 \times 1$  畳み込みを行い, チャンネル方向のみ計算を行う。

BCI のモデルは大きくなる程, より多くのデータまたはより多くのモデルの正規化 (またはその両方) が必要となり, 学習すべきハイパーパラメータが増え, ハイパーパラメータの誤指定によるモデル性能のばらつきが大きくなる可能性がある。EEGNet は Depthwise 畳み込みや Separable 畳み込みを取り入れることにパラメータ数を通常の畳み込みに対して減らすことができ, オーバーフィッティングを起こしにくいモデルとなると考えられている。また, このような構造は意味のある特徴量を抽出可能であることが示されて

Table 1 EEGNet architecture

Block	Layer	filter	Output	Activation	Dropout
1	input				
	Reshape				
	Conv2D	16	(16,122,401)	Linear	
	BatchNorm		(16,122,401)		
2	DepthwiseConv2D		(2*16,1,401)	Linear	
	BatchNorm		(2*16,1,401)		
	Activation		(2*16,1,401)	ELU	
	AveragePool2D		(2*16,1,100)		
	Dropout		(2*16,1,100)		p=0.25
3	SeparableConv2D	32	(32,1,100)	Linear	
	BatchNorm		(32,1,100)		
	Activation		(32,1,100)	ELU	
	AveragePool2D		(32,1,12)		
	Dropout		(32,1,12)		p=0.25
	Flatten		(32,1,12)		
Classification	Dense		3	Softmax	

いる [2]. つまり, EEGNet は神経生理学的信号から解釈可能な抽象的特徴を抽出する上, 少ないデータセットで信頼性の高いモデルを構築することが可能である.

### 3.2 識別手法

EEGNet を用いて音声想起の識別を行い, 音声想起時の EEGNet の有用性を調べる. このとき, 3つの音声はランダムで流れるため, データ数は被験者ごとに異なるが1人あたりの平均収録データ数は322である. 収録データの80%を学習データ, 20%をテストデータとして5 fold cross validation を行い, 識別精度の平均を算出した. また, データを見て明らかに挙動がおかしいチャンネルの脳磁界データを除外し, そのデータに対して白色化を行った. 本実験では, ランダム雑音を軽減するために10回加算を行なった波形を用いている. 加算波形は単一試行波形のデータ数と同数になるように作成した. パラメータを Table 1 に示す. ネットワークの損失関数には交差エントロピー誤差, 勾配降下アルゴリズムは Adam[6] を用い, バッチサイズは32, エポック

数は50として学習を行った. また, 音声想起識別の実装には MNE-Python を用いた [7].

## 4 実験結果と考察

単一試行波形・加算波形に対して識別実験を行い, 音声想起時の被験者ごとの識別精度を Fig.3 に示す.

単一試行波形について, 平均識別率 37.3% と, chance rate (33.3%) よりも上回っている. また, 被験者2は識別率が一番低い34.8%であるが, これも chance rate を上回っている. このことから EEGNet は音声想起のタスクにも適用可能であると考えられる.

10回加算波形について, 全被験者で単一試行よりも精度が上がり, ランダム雑音を減少できていることがわかる. しかし, 単一試行波形よりも10回加算波形の被験者によつての精度のばらつきが大きい. これは, 文字呈示のトリガが入力されるとその立ち上がりに同期して計測データを加算しているため, 想起のタイミングがずれていることが考えられる.

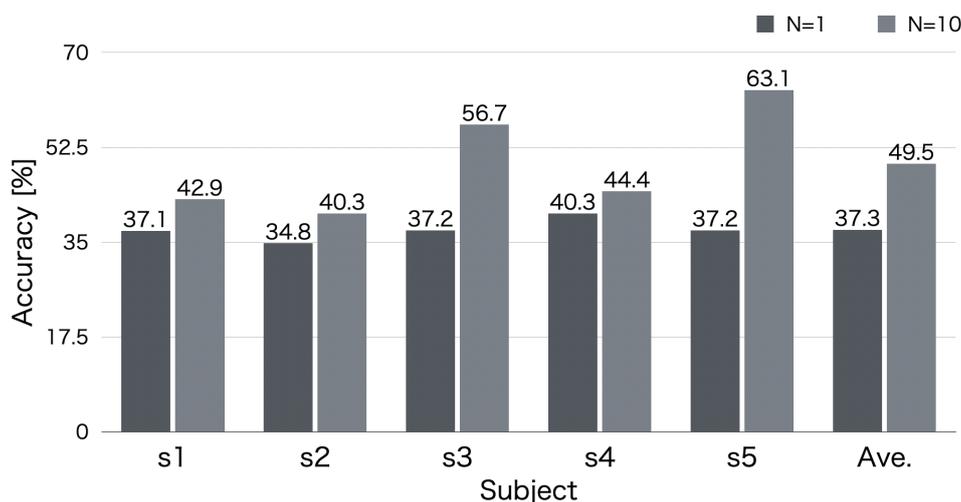


Fig. 3 Discrimination accuracy of the imagined sounds.

## 5 まとめ

脳磁界データにおける音声想起のデータの少なさを克服するため、様々なBCIタスクにおいて適用可能なEEGNetを用いた音声想起を提案した。実データを用いた実験を行うことによって、EEGNetは音声想起に適用可能であることを示した。

今後の展望としては、想起のタイミングの同定や実データの処理を行い、よりの確に特徴を捉えた識別を行うモデルを検討し実用上十分な識別精度を実現したい。

**謝辞** 本研究の一部は、JSPS 科研費 JP22K18626, JP21H05596 の助成を受けたものである。

## 参考文献

- [1] R. Fazel-Rezai et al., “P300 brain computer interface : current challenges and emerging trends,” *Frontiers in Neuroengineering*, pp. 1-15, 2012.
- [2] V. Lawhern et al., “EEGNet: a compact convolutional network for EEG-based brain-computer interfaces,” *Journal of Neural Engineering* 15(5), 1741-2552, 2016.
- [3] 矢野ら, “Convolutional Neural Networks を用いた音声想起時の脳磁界データにおける識

別的特徴量の検討,” 日本音響学会 2017 年秋季研究発表会, pp. 507-510, 2020.

- [4] S. Uzawa et al., “Spatiotemporal Properties of Magnetic Fields Induced by Auditory Speech Sound Imagery and Perception,” *IEEE EMBC2017*, pp. 2542-2545.
- [5] R. Shi et al., “Categorizing objects from MEG signals using EEGNet,” *Cognitive Neurodynamics* 2022, 16:365-377.
- [6] P. Diederik et al., “Adam: A Method for Stochastic Optimization,” *arXiv preprint arXiv:1412.6980*.
- [7] A. Gramfort et al., “MEG and EEG data analysis with MNE-Python,” *Frontiers in Neuroscience*, 7, 267, 2013.