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 *

☆矢野彩緒里, 高島遼一, 滝口哲也, 有木康雄 (神戸大),
添田喜治 (産総研), 中川誠司 (千葉大/産総研)

1 はじめに

近年，身体障害者のコミュニケーション支援を目的として，音声認識や行動認識が用いられている．しかし，発話や身振りを使ったコミュニケーション手段を用いることが困難であることからそのような技術を活用できない身体障害者も存在する．そこで，脳活動を用いて機械制御やコミュニケーションを行う，ブレイン・コンピュータ・インターフェイス (BCI) の開発に向けた研究が盛んになっている．

これまでの BCI は，P300 スペラー型 [1] が多く，これは注意を向けた低頻度刺激に対する誘発反応 (P300) を用いて限定的に意思伝達を行うことができる．しかし，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であり，自由で汎用的なコミュニケーションを行うことはできない．そこで我々は，利用者の意思をより自由に理解するような BCI を開発することを目的として，脳磁界データから想起音声の識別を試みた．

我々はこれまで，脳磁界計測データを用いて，音声聴取時との比較から音声想起時の時空間特性の検証を行ってきた．想起音声のエンベロープと想起により誘発された加算平均脳磁界反応の相関を調べたところ，想起時には音声聴取時と同レベルの相関は得られなかったものの, 時間波形上で聴覚野周辺の緩やかな活動が認められた [2]．また，想起音声の単語識別において，半数の被験者において音声想起時と聴取時に識別制度が高いチャネルが一致し，空間的に類似している可能性が見られた [3].

想起音声の識別率を向上させることを目的とした試みでは，脳磁界データのチャネル・周波数・時間における多次元情報を損なわないよう，テンソル分解により低次元な特徴量を抽出し，学習モデルの評価を行った [4]． さらに，空間的な特徴を損なわない識別についても検討を行ったが，実用上十分な識別精度が得られたとは言い難い [5].

識別制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，特徴量に対し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．目的とする反応が微弱であり，膨大な特徴量を持つ脳磁界データを用いた識別において，特徴を的確に捉えることは極めて重要である．

本稿では，脳磁界データから抽出したウェーブレット特徴量から，Squeeze-and-Excitation Networks[6] により識別に有用なチャネル選択を行い，識別に必要な特徴のみを抽出することを目的とした Convolutional Neural Networks (CNN) モデルを実装し，従来の識別手法との比較を行った．

2 脳磁界計測及び特徴量抽出

2.1 脳磁界計測

8 名の右利き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．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 回目の文字呈示のタイミングで聴取した音声を聞こえたままに想起するよう求めた．1 試行の概要を Fig. 1 に示す [2].

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 800 ms であり，実験に用いた脳磁界データの時間は，想起開始時から 800 ms 間とした．

脳磁界計測には，122 ch 全頭型脳磁界計測シス

* Discriminative features in brain magnetic fields during auditory speech sound imagery using convolutional neural networks YANO, Saori, TAKASHIMA, Ryoichi, TAKIGUCHI, Tetsuya, ARIKI, Yasuo (Kobe Univ.), SOETA, Yoshiharu (AIST), NAKAGAWA, Seiji (Chiba Univ./AIST).

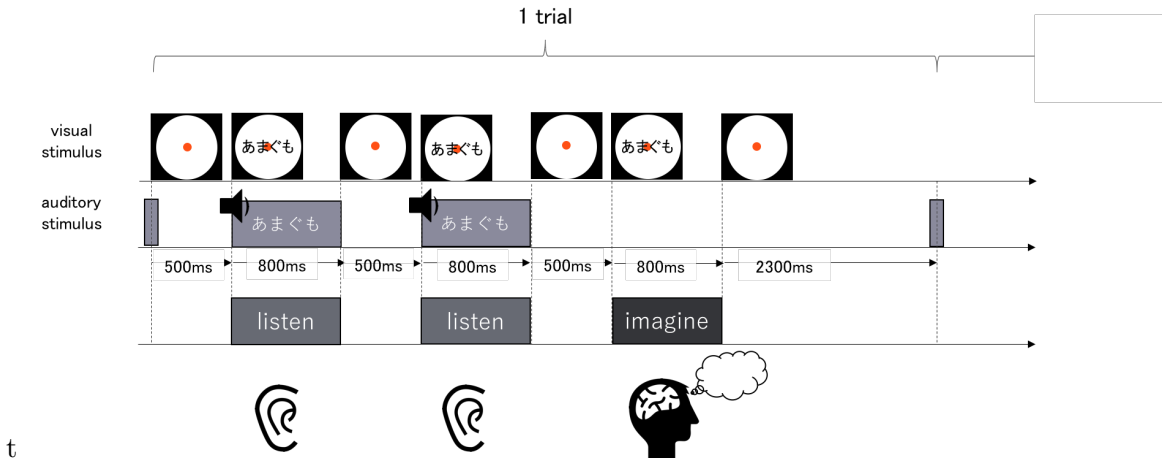

Fig. 1　Schematic diagram of the task.

テム (Neuromag - 122^{TM}: Neuromag, Ltd.) を用いた．計測した脳磁界データは 0.03-100 Hz のアナログフィルタを適用した後，サンプリング周波数 400 Hz で A/D 変換を行った．同時に被験者 1〜 3 については眼電図 (Electrooculogram : EOG) の測定も行い， EOG が 5 mV を超えた際の脳磁界データは学習・テストデータから除外した．

2.2　特徴量抽出

本実験では単一試行波形に対して独立成分分析 (Independent Component Analysis : ICA) により眼電除去を行った．左右側頭部聴覚野を覆う計 36 チャネルに対して，文字刺激提示開始時から 800ms 間の単一試行波形または加算波形に対して，連続ウェーブレット変換 (Continuous Wavelet Transform : CWT) による時間周波数特徴量の抽出を行った．CWT 関数 W は以下の式 (1) で示される．

$$W(a,b) = \frac{1}{\sqrt{a}} \int x(t)\psi(\frac{t-b}{a})dt \qquad (1)$$

ここで，$x(t)$ は脳磁界の時系列波形である．$\psi(t)$ はマザーウェーブレットであり，本実験では Morlet ウェーブレットを用いた．a はスケール，b は時間シフトを表すマザーウェーブレットのパラメータである．

CWT 関数 W には，1 チャネルごとに周波数方向に 1- 25 Hz の 25 次元，時間フレームとして 320 次元が得られる．本実験では特徴量削減のため，フレーム長 100ms，フレームシフト 50ms とし，時間フレームを 15 次元に圧縮した．

3　識別モデル

3.1　モデル概要

識別のモデルには多チャネル時間周波数特徴量に対し二次元の畳み込みを行う Convolutional Neural Networks (CNN) を用いた．時間方向や周波数方向の特徴を維持したまま高次元特徴量を捉えることが目的である．また，音声想起において，多チャネルの情報から必要な情報を取り出すため，Squeeze-and-Excitation Networks を組み込んだ．モデルの概要を Fig. 2 に示す．

まず，CNN モデルは入力を 36 チャネルのウェーブレット特徴量とし，2 層の畳み込み層と 2 層のマックスプーリング層，2 層の全結合層から構成される．畳み込み層のカーネルフィルタ数は一層目は 64，二層目は 128 であり，カーネルフィルタサイズは一層目は 3×3，二層目は 2×2 であり，ストライドは 1×1 である．また，マックスプーリングの領域は 2×2 とした．

全結合層のノード数はそれぞれ 128，3 であり，出力には (“あまぐも”，“いべんと”，“うらない”) の 3 単語にそれぞれ対応する 1 hot ベクトルが得られる．活性化関数には relu，出力層にのみ softmax を用い，損失関数にはクロスエントロピーを，勾配降下アルゴリズムは adagrad[7] を用いた．

3.2　Squeeze-and-Excitation Networks

本研究の提案手法において，CNN モデルの入力層と畳み込み層の間に，Squeeze-and-Excitation Networks (SENets) を組み込んだ．これにより，入力特徴量に対し各チャネルを均等に

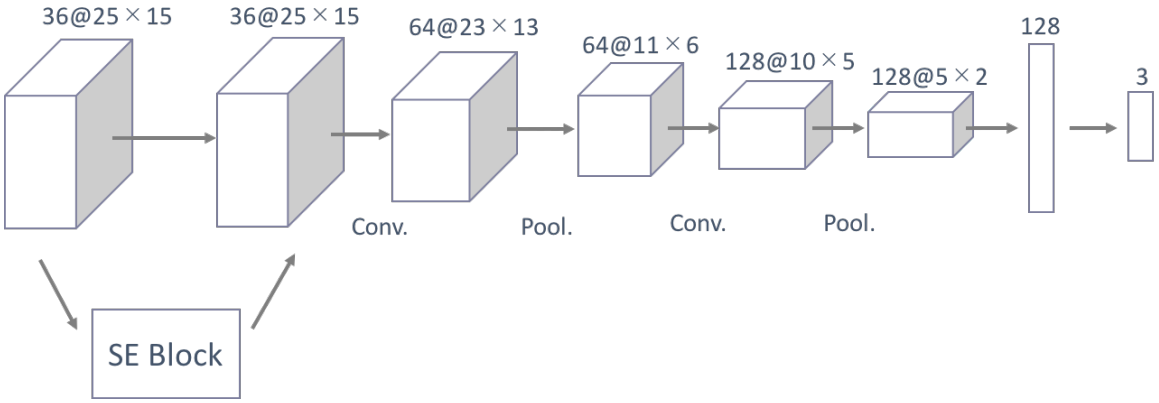

Fig. 2　Proposed CNN model.

出力せず，適応的に重みをかけることが可能となる．SENets の構造を Fig. 3 に示す．

C をチャンネル数，F を周波数方向の次元数，T を時間方向の次元数とする．まず，入力特徴量の各チャネルに対し，平均値を抽出し各チャネルの特徴量とし，２層からなる全結合層への入力とする．出力として得られる値は各チャネルの固有の重みとして，元の入力特徴量に乗算される．これによりチャネルごとの重み付けを実現させる．本研究では，全結合層のボトルネックを形成する圧縮率 r を 2 とする．

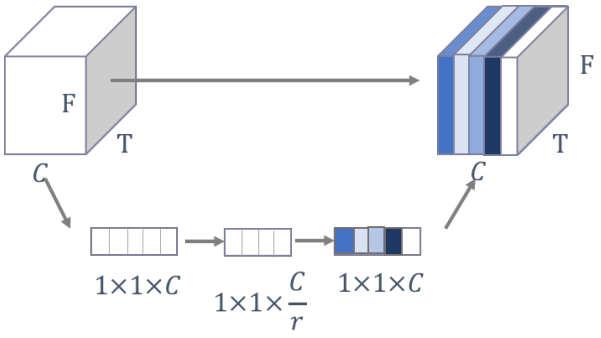

Fig. 3　Squeeze-and-Excitation Networks.

4　実験

本稿では，提案手法である CNN+SENets と，CNN，SVM において識別を行い，精度を比較した．SVM については，次元削減のため，主成分分析 (Principal Component Analysis : PCA) を用いて 1 チャネルにつき 10 次元に次元削減を行なったものを識別器の入力特徴量としている．

単一試行波形，10 回加算波形のそれぞれを入力波形とし，上記の実験を行った．それぞれの波形における被験者ごとの識別率を Table 1 に示す．なお，表中の N は波形の加算回数を示している．

単一試行波形について，提案手法は平均識別率 37.9% と，CNN，SVM より僅かだが高い値を示した．また，被験者 6 において，CNN，SVM では識別率が chance rate に近く学習がうまく行えていないが，提案手法では 40.0 % の精度を示した．一方，被験者 4 については CNN を用いた識別が chance rate であるのに対し，SVM が 39.6% の識別率を示した．

10 回加算波形について，提案手法の識別率は CNN を 6.4%，SVM を 8.7% 上回る 59.3% を示し，識別精度を大きくあげることができた．特に，被験者 1 に関しては CNN より 15.5%, 被験者 4 については SVM より 15.9% 高い識別率を示し，他の識別手法より 15% 以上高い識別率を得ることができた．また，他の手法より極端に識別率が低くなることもなく，安定した識別精度で分類することができた．

5　考察

単一試行波形においては，提案手法の有効性を明確に示すことができなかった．被験者 4 において SVM の識別精度が CNN の識別精度を大きく上回った理由としては，SVM では PCA により次元数を減らしたことによって識別が容易になったことに対し，CNN においては高次元特徴量から適切に特徴を捉えることができなかったためであると考える．一方，同被験者において 10 回加算波形の CNN+SENets の識別率は他手法の識別率を大きく上回っている．以上より，音声想起に伴う脳磁界信号が微弱であるが，波形

Table 1　Classification accuracy [%].

N=1	Sub.1	Sub. 2	Sub. 3	Sub. 4	Sub. 5	Sub. 6	Sub. 7	Sub. 8	Ave.
CNN + SENets	41.5	44.1	36.8	33.3	40.9	40.0	33.9	32.4	**37.9**
CNN	44.3	35.6	34.7	33.3	43.9	33.9	38.7	36.8	37.7
SVM	43.0	35.6	33.6	39.6	30.3	33.8	36.5	39.7	36.5
N=10	Sub.1	Sub. 2	Sub. 3	Sub. 4	Sub. 5	Sub. 6	Sub. 7	Sub. 8	Ave.
CNN + SENets	**70.9**	52.5	40	**63.5**	68.2	58.5	66.1	54.4	**59.3**
CNN	54.4	54.2	45.3	39.7	71.2	41.5	59.7	57.4	52.9
SVM	37.8	49.1	31.6	47.6	57.4	66.2	61.0	54.4	50.6

を同期加算することで雑音除去が行われ目的の信号を取り出すことができており，その波形に対しチャネル選択が的確に行われていると認識できる．このような理由から，単一試行波形においても，雑音除去を適切に行い目的の信号を抽出することができれば，SENets によって識別精度を向上できるのではないかと考える．

また，10 回加算波形において，CNN+SENets による識別率が他手法より極端に高い被験者は見られるが，極端に低い被験者は見られなかった．以上においても，SENets が雑音成分を除去した状態の波形に有効に作用していると考えることができる．

6　おわりに

本稿では，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，単一試行波形の識別精度を僅かながら向上させ，また，加算波形の識別精度を大きく向上させることができた．

今後の展望としては，チャネル方向だけでなく，時間や周波数帯域にも重み付けを行い，より的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 JP18K19820 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] S. Uzawa *et al.*, “Spatiotemporal Properties of Magnetic Fields Induced by Auditory Speech Sound Imagery and Perception,” IEEE EMBC2017, pp. 2542-2545.

[3] 矢野ら, “脳磁界データによる音声の識別-想起時と聴取時の比較-,” 日本音響学会 2019 年秋季研究発表会, pp. 647-650.

[4] 宇澤ら, “脳磁界データによる想起音声の識別-次元数削減による精度向上の検討-,” 日本音響学会 2017 年 秋季研究発表会, pp. 337-340.

[5] 矢野ら, “脳磁界データの空間的特徴を考慮した想起音声の識別,” 日本音響学会 2018 年 秋季研究発表会, pp. 337-340.

[6] J. Hu *et al.*, “Squeeze-and-Excitation Networks,” CVPR, pp. 7132-7141, 2018.

[7] J. Duchi *et al.*, “Adaptive Subgradient Methods for Online Learning and Stochastic Optimization,” Journal of Machine Learning Research 12, pp. 2121-2159, 2011.