

脳磁界データの空間的特徴を考慮した想起音声の識別*

☆矢野彩緒里, 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

現代の社会では、音声認識や行動認識を用いた機械制御・意思伝達技術が様々な場面で用いられている。しかし、会話や身振りでのコミュニケーションが困難な障害者支援はそのような技術を用いてうまく利用することができない。そこで、彼らがより便利な生活を送れるよう支援するために、脳活動を用いて機械制御や意思伝達をおこなう Brain Computer Interface (BCI) の利用が期待される。

これまでの 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おこなう “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。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おこなってきた。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、脳磁界データのチャンネル・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 [3]。

本研究は、これまでの特徴量に加え、チャンネル同士の空間的な相関を考慮する。EEG データの分類研究では、チャンネル同士の位置関係をグラフとして用いる手法 [4] や、チャンネル位置に基づいた画像情報として表現する手法 [5] の提案がされている。

本稿では、脳磁界データを (2次元画像) × (周波数) の3次元の特徴量として表現することで、空間情報・周波数情報を維持したまま想起音声の分類をおこなった。

2 脳磁界データの計測

3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”) を用いて実験をおこなった。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回呈示をおこなった。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、解析対象の脳磁界データの解析時間は想起開始の合図から 1000 ms とした。

脳磁界計測には、122 ch 全頭型脳磁界計測システム (Neuromag - 122TM: 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 変換をおこなった。同時に眼電図 (Electrooculogram : EOG) の測定もおこない、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。

3 特徴量抽出と分類モデル

3.1 特徴量抽出

脳磁界計測で得られる波形の加算平均を Fig. 2 に示す。

加算平均をおこなうことで、脳磁界データのノイズ成分を低減させることができる。本実験では単一試行波形を用いるため、環境雑音や筋電による雑音を多く含む。単一試行により得られる波形を Fig. 3 に示す。

* Auditory imagery classification based on spatial features of brain magnetic fields, YANO, Saori, TAKIGUCHI, Tetsuya, ARIKI, Yasuo (Kobe Univ.), SOETA, Yoshiharu (AIST), NAKAGAWA, Seiji (Chiba Univ./AIST).

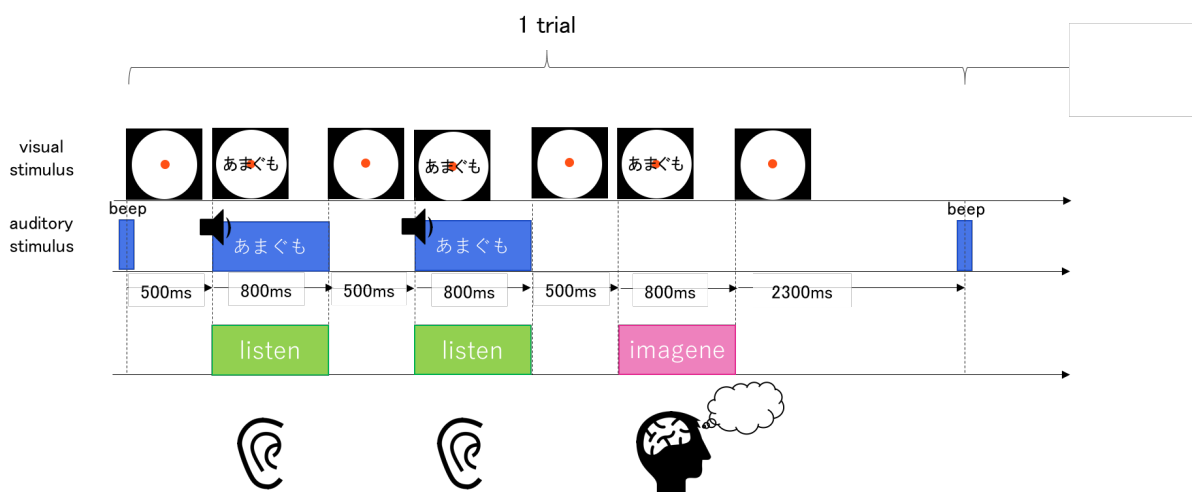


Fig. 1 Schematic diagram of the task.

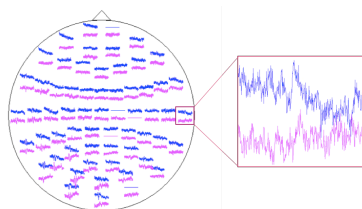


Fig. 2 Averaged waveform during auditory imagery.

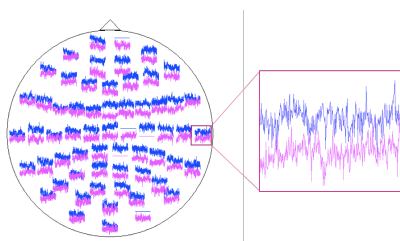


Fig. 3 Single-trial waveform during auditory imagery.

各単語を想起した際に右半球側頭部において測定された単一試行波形と、パワースペクトル密度 (PSD) を Fig. 4 に示す。

特定の時間・周波数における PSD を、チャンネル位置を考慮して 2 次元画像上にマッピングすると、Fig. 5 のような画像データが得られる。ここで、この 2 次元画像は頭部を上から見たもので、画像上方は頭部の前方を示す。こうして作成された画像データを特徴量として扱う。

次に、上で述べた特徴量を抽出する手順を Fig. 6 に示す。脳磁界測定では、チャンネルごとに

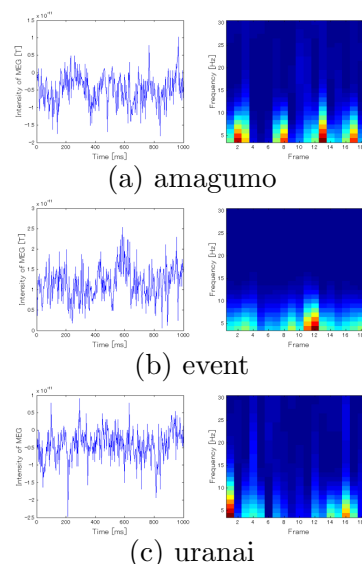


Fig. 4 Single-trial waveform and PSD.

時系列の波形データが得られる。得られたデータより PSD を算出し、時間・周波数特徴量とする。次に、チャンネル位置から推測される 2 次元平面上の座標に対応する値をそのチャンネルの PSD とし、マッピングする。チャンネル位置に基づく散布データを、線形補間によって補間し、画像情報とする。こうして得られた画像データに対し、取りうる値が 0 ~ 1 となるよう正規化をおこない、入力特徴量とする。

1 試行から得られる特徴量の次元数は、(2 次元画像) × (周波数) × (時間) となる。ここで、フレーム長を 100 [ms]、フレームシフトを 50 [ms]、周波数帯域を 4 - 30 [Hz]、周波数間隔を 1 [Hz]、画像サイズを 11 × 11 [px] とする。

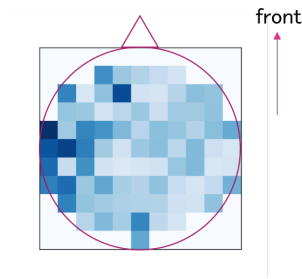


Fig. 5 Features based on the spatial correlation.

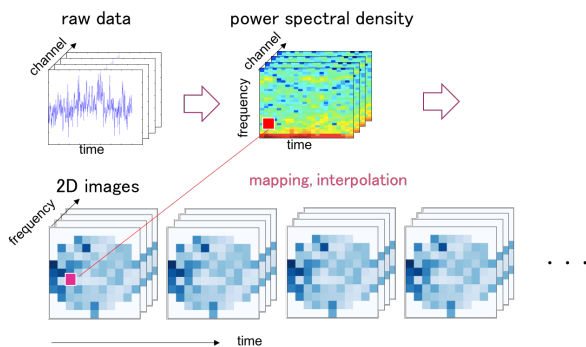


Fig. 6 Feature extraction method.

3.2 分類モデル

分類の DNN モデルには、画像情報の特徴を取り出す Convolutional Neural Network (CNN) と、連続時間情報を記憶し回帰する Long-Short Time Memory (LSTM) を組み合わせた CNN-LSTM モデル [6] を用いた。モデルの概要を Fig. 7 に示す。

まず、CNN において、(2次元画像) × (周波数) の 3次元特徴量を入力として、出力にはそれらから抽出される特徴を得る。ここで、CNN モデルは 1層の畳み込みレイヤと 1層のプーリングレイヤから構成される。チャンネル同士の空間的相関だけでなく、隣り合う周波数帯域の相関が強いと予想されることも考慮して、畳み込み層、プーリング層はそれぞれ 3次元を採用した。CNN で抽出された特徴量は、1次元に平滑化される。フレームごとの CNN の出力を、LSTM の入力とし、LSTM の出力は 1層の全結合層に入力され、出力には各クラスに対応する 1 hot ベクトルを得る。本モデルは、CNN と LSTM まで連続しており、別々に学習するようにはなっておらず、勾配の更新は CNN-LSTM モデルを通しておこなわれる。

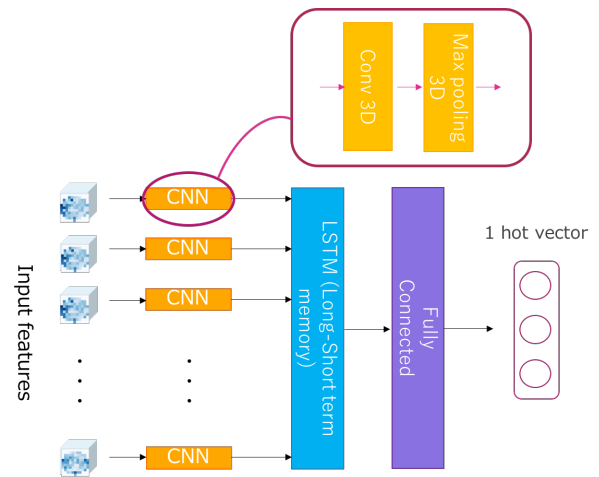


Fig. 7 Proposed DNN model.

畳み込み層のカーネルフィルタ数は 16、カーネルフィルタサイズは $3 \times 3 \times 3$ とする。また、LSTM 層、全結合層はともにノード数を 100 とし、出力には 3次元の 1 hot ベクトルが得られる。出力により入力特徴量は 3 クラスに分類され、(“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。LSTM 層、全結合層の活性化関数はそれぞれ relu, softmax とし、全結合層は 50% をドロップアウトして学習をおこなった。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 adam[7] を用いた。

4 実験

4.1 実験条件

脳磁界計測により得られたデータより特徴量抽出をおこない、上記モデルにおいて分類をおこなった。眼電が多く混入したデータを破棄したため、学習・テストデータ数は各被験者によって異なる。被験者ごとの学習・テストデータ数を Table 1 に示す。

Table 1 Number of data.

| Subject | Sub. 1 | Sub. 2 | Sub. 3 |
|---------------|--------|--------|--------|
| training data | 240 | 224 | 370 |
| test data | 61 | 57 | 93 |

モデルの評価は 5 fold cross validation をおこない、正解率の平均によっておこなう。また、データを分割する際、計測順序に起因したデータの偏りを防ぐため、分割されたデータ群をそ

それぞれ $D_k (k = 0, 1, 2, 3, 4)$ とすると, i 回目の試行について $i \equiv k(\text{mod}.5)$ とするとき, データ群 D_k に属するように決める.

また, 本稿で比較手法として, 画像上にマッピングをおこなわない時間周波数特徴量を, 提案手法の CNN モデルの層を除いた LSTM モデルを用いて分類した.

4.2 結果と考察

実験結果を Table 2 に示す.

Table 2 Classification accuracy [%].

| model | LSTM | Proposed |
|--------|------|----------|
| Sub. 1 | 36.9 | 39.4 |
| Sub. 2 | 33.2 | 32.4 |
| Sub. 3 | 31.5 | 33.7 |

被験者 1 において, 提案手法のモデルは LSTM モデルの精度を 2.5 % 上回った. 被験者 1 に対しては提案手法が有効にはたらいっていると言える. しかし識別モデルとしての精度は低く, さらに精度を向上させる必要がある.

被験者 2 においては提案手法の識別率の方が低く, 被験者 3 は提案手法のモデルが 2.2 % 上回っているが, chance level 付近にとどまった.

これは, 提案モデルにより特徴がうまく取り出せていないことを示している. 理由としては, 脳磁界波形に環境雑音や筋電による雑音が大きく影響していること, データ数が少ないことが上げられる.

脳磁界データは一般的に波形を加算することによりノイズキャンセリングがおこなわれる. 雑音による影響を小さくするための手段として, 加算波形を用いて分類をおこなったモデルを転移学習させることが考えられる. また, 加算波形を用いた分類精度は単一試行波形に比べ高いと予測され, 提案手法と従来手法の比較がより明確におこなわれると考える.

また, 特徴量に対しデータ数が少ないことの改善としては, 特徴量の次元削減, データ数の増加の双方によりおこなうことができる. 脳次回データの測定の際には被験者を長時間拘束する必要があり, 集中力の欠如がデータの質に影響することを考慮すると, データ数を増やすためには同じ実験時間の中でより多くのデータを得られるようなタスクの改善が求められる. また,

特徴量抽出において, 脳磁界測定 of 脳磁界データの空間的特徴を維持したまま次元削減をおこなう方法の検討が必要である.

5 おわりに

本稿では, 脳磁界データから想起音声の分類をおこなうタスクにおいて, 脳活動の空間的特徴を取り入れるため, 各チャンネルの PSD を 2 次元にマッピングし, 画像情報として扱った. しかし, 脳磁界データの各試行波形は誘発反応以外の雑音成分を多く含むこと, 特徴量の次元に対して十分な数のデータ測定が行えていないことから, 認識率は chance level を少し上回る程度に留まってしまった.

今後は脳機能計測実験をさらにおこない, 一回の実験において得られるデータ数を増やすことで, 認識精度の向上を図りたい. また, 単一試行波形における雑音除去の手段を検討する必要がある.

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] 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.
- [4] S. Jang *et al.*, “EEG-based Video Identification using Graph Signal Modeling and Graph Convolutional Neural Network,” *IEEE ICASSP2018*, pp. 3066-3070.
- [5] S. Kuanar *et al.*, “Cognitive Analysis of Working Memory Load from EEG, by a Deep Recurrent Neural Network,” *IEEE ICASSP2018*, pp. 2576-2580.
- [6] J. Donahue *et al.*, “Long-Term Recurrent Convolutional Networks for Visual Recognition and Description,” *IEEE CVPR2015*, pp. 2625-2634.
- [7] D. P. Kingma *et al.*, “Adam : a method for stochastic optimization,” *International Conference on Learning Representations 2015*.