

音声想起時の脳波における想起区間の推定

鈴木大裕¹ 入部百合絵¹

¹愛知県立大学 大学院情報科学研究科

im222010@cis.aichi-pu.ac.jp iribe@ist.aichi-pu.ac.jp

概要

音声想起時脳波から想起言語を識別する BCI 研究が始まっている。音声想起識別では想起区間を検出せず、ノイズを含む区間を想起対象データとして扱うため識別精度が低下する傾向にある。そこで想起言語の識別精度向上を目標に、本稿では音声想起時脳波における想起区間の推定について報告する。本報告では、定 Q 変換によるスペクトログラムをサンプル加算することで想起区間を得て、想起区間を基に 3DCNN で想起区間推定を行った。その結果、加算スペクトログラムで得た反応を想起区間として 3DCNN に入力することで、約 80% の精度で想起区間推定に成功した。また、データを加算することで約 90% の推定精度を得た。

1 はじめに

BCI (Brain Computer Interface) では、神経疾患を持つ方の円滑なコミュニケーションを実現するため、脳波から音声言語情報を抽出する研究が行われている[1]。これらの研究では、発話時の脳波を対象とするものと、音声想起 (speech imagery) 時の脳波を用いるものが存在する。音声想起では相手に話しかけるイメージで言語を想起するのみとし、原則、発話器官は動かさない。本研究では、神経疾患を持つ方々に寄与するため、音声想起時の脳波を対象とする。

発話音声からの音声認識では、発話区間の推定が識別精度に大きく影響する。音声想起でも同様のことが言えるだろう。しかし、音声想起を用いて想起言語を推定する研究では、脳波から想起区間を特定することが困難であるため、「無想起区間を多く含む脳波データ」を対象に想起言語を識別せざるを得ない。その結果、無想起区間が干渉することで、想起言語の識別精度が劣化する傾向にある。

本研究では、音声想起区間を特定することで、音声想起における言語推定の精度向上を目指す。本稿では、音声想起区間の検出手法について提案する。

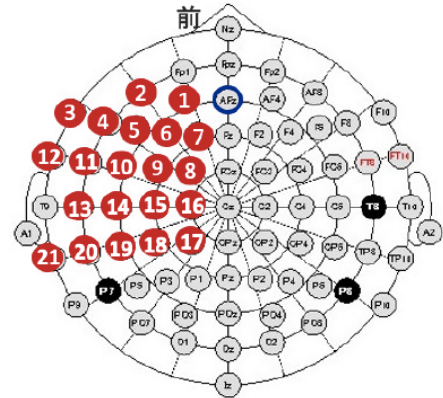


図 1 左脳側 21 電極の配置

先行研究[2]では、CSP (Common Spatial Pattern) を特徴量とし混合ガウスモデル (Gaussian Mixture Model) に音声想起と無想起の 2 クラス分類した結果、約 80% の精度を得たが、音声想起区間の推定までは行えていない。本報告では、定 Q 変換[3]を用いた加算スペクトログラムによる想起反応の調査を行い、その想起反応を用いて 3DCNN (3 Dimensional Convolutional Neural Network) [4] による想起区間の推定を行う。また、先行研究の手法を用いた区間推定と比較する。

2 音声想起時脳波の収録

インフォームド・コンセントを得た右利き成人男性 5 名 (A, B, C, D, E) を対象に、日本語母音 /a/ の脳波を収録した。脳波計測には g.tec 社の g.HIAMP を使用し、サンプリング周波数 512Hz で計測した。電極は拡張 10-20 法に従い、音声想起活動があるとされる左脳側に 21 電極 (1; AF3, 2; AF7, 3; F9, 4; F7, 5; F5, 6; F3, 7; F1, 8; FC1, 9; FC3, 10; FC5, 11; FT7, 12; FT9, 13; T7, 14; C5, 15; C3, 16; C1, 17; CP1, 18; CP3, 19; CP5, 20; TP7, 21; TP9) を配置し (図 1; 赤丸は電極番号)、基準電極を左の耳朶、アース電極を AFz (図 1; 青丸) とした。

脳波研究では収録実験をする際、プロトコルを用いる。本研究では、図 2 のプロトコルを作成した。同じ単語を連続で想起した場合、想起脳波が脳内で

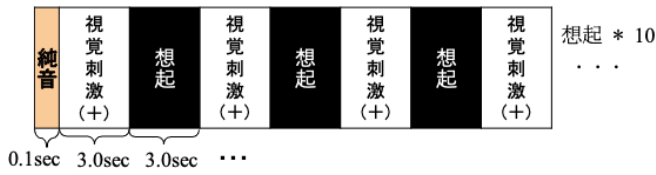


図 2 視覚刺激による収録プロトコル

時間的に重畳され強い反応となり、識別率が向上する[5]。そこで本収録実験では、想起反応を容易に取得可能にするため、母音/a/のみをプロトコル内で10回想起した。また、事前に実施した実験では純音による強い聴覚反応が観察されたため、視覚刺激によるプロトコルとした。

収録実験では、電磁シールドルーム内の椅子に被験者が座り、視覚刺激となるディスプレイを見せながら収録を実施した。想起の時間間隔は3.0sec設けられているが、被験者には画面の十字マークが消えたら即座に/a/と1回だけ想起するように伝えた。また、/a/の文字は思い浮かべないこと、音声を想像すること（相手に話しかけるイメージ）、瞬きはノイズとして重畳されてしまうため想起中は瞬きを極力しないこと（可能な限り十字マーク表示の際に瞬きをする）の3点を指示した。

以上の実験を被験者5名に対して、プロトコル1セット10回想起を20セット行い、日本語母音/a/の音声想起を被験者毎に200サンプル収集した。

3 提案手法

本報告で提案する手法は、前処理、特徴量抽出、想起区間推定の順で行う。

3.1 前処理

脳波データには脳内アーチファクトや電源ノイズの影響があるため前処理を行う。前処理では、直流成分と電極間バイアス、電源ノイズの除去、21電極の加算、逆位相キャンセリングを実施する。

3.1.1 直流成分除去

本研究で使用するg.HIAMPはDCアンプのため、EEG信号にはオフセットが混在する。直流成分が含まれると、のちの周波数解析で正しい結果が得られない場合があるため直流成分を除去する。本研究では除去方法として、時間信号 $x(t)$ に対して前後50msの区間を平均化し、 $x(t)$ から引き去る。計算式を式(1)に示す。 $x(t)$ は時間信号、 t は時間(ms)、 \bar{x} は $t-50 \sim t+50$ までの平均値とする。

$$y(t) = x(t) - \bar{x} \quad (1)$$

3.1.2 電極間バイアス除去

各電極で計測される電位は、基準電極と各測定電極の差分として得ているため、電極間で異なるバイアスを持つ。そのため、電極間のバイアスを除去する必要がある。全電極の信号を平均したものを基準値とし、各電極から引く処理を入れることで、電極間のバイアスを平準化する。計算式を式(2)に示す。 ch は電極番号、 t は時間、 $x(ch, t)$ は各電極の時間信号、 $\overline{x(t)}$ は全電極を平均した時間信号である。

$$y(ch, t) = x(ch, t) - \overline{x(t)} \quad (2)$$

3.1.3 電源ノイズ除去

収録したデータには、電源ノイズが含まれる。電源ノイズは、脳波波形の観測を困難にする。そのため、必ず除去する必要がある。

電源ノイズは急峻なノイズである。そこで、特定の狭い周波数帯のみを減衰するノッチフィルタによりノイズ除去を行う。本研究の収録場所は愛知県であるため、60Hzとその高調波120, 180, 240Hzに対し、ノッチフィルタを適用することで電源ノイズを除去する。

3.1.4 21電極の加算

21電極加算は、音声想起の特徴を捉えるための前処理である。皮質脳波の研究[6]では音声発話と音声想起、聴取のデータを収集し、振幅特性を分析しているが、音声想起反応のある電極にばらつきが生じていた。一方、本研究ではキャップ式脳波計を用いるため、被験者毎に反応電極が異なる事象が発生する。以上の事由により、本研究では21電極を加算することとした。計算式を式(3)に示す。 CH は総電極数、 ch は電極、 t は時間、 $x(ch, t)$ は各電極の時間信号である。

$$y(t) = \sum_{ch=1}^{CH} x(ch, t) \quad (3)$$

3.1.5 逆位相キャンセリング

逆位相キャンセリングは、位相スペクトルを反転させることでノイズを減衰する技術である。ヘッドホンなどに利用されている。

21電極を加算することで、全電極に生じる音声想起反応以外の反応（ノイズ）が要因となり、波形にゆがみが観測された。そのため、一定以上の強さをもつスペクトルを位相反転することで、波形の歪みを減衰させた。これにより、体動のような強いノイズを除去する。

3.2 音声想起の特徴量抽出

特徴量抽出では周波数解析を行う。言語表象は high- γ 帯域(60-180Hz)に現れると報告されている[7]。言語表象は音声想起反応の一部と考えられるため、本研究では high- γ 帯域のような高周波帯域も対象に、低周波帯域から広域に亘り周波数分析を行う。そのため定 Q 変換を行い、さらに定常雑音を除去するためのスペクトルサブトラクションを施した。

3.2.1 定 Q 変換 [3]

短時間フーリエ変換では、全ての周波数帯で同じ窓長を用いることで周波数変換をする。それにより、時間分解能を一定に保ち、高周波数帯では高い周波数分解能を持つ。しかし、低周波数帯では十分な周波数分解能を持つことができない。そこで、低周波帯域と高周波帯域を一定の分解能で分析するため、定 Q 変換を用いる。

定 Q 変換とは、中心周波数ごとに窓長を変えることを実現した変換法である。中心周波数が高い時は窓長を狭くし、低い時は窓長を広くすることにより、全ての周波数帯域で一様な周波数分解能を得られる。計算式を式(4)に示す。なお、 n は時間サンプル、 x は時間信号、 ω は窓関数、 k は対数周波数のビン数、 $N[k]$ は窓長の変化、指数関数は中心周波数に対するフィルタを表す。

$$y(k) = \frac{1}{N(k)} \sum_{n=0}^{N(k)-1} \omega(n, k) x(n) e^{-\frac{2j\pi Qn}{N(k)}} \quad (4)$$

3.2.2 スペクトルサブトラクション

収録した脳波データには呼吸や心臓の鼓動、脳内で定常的に発生している信号など、脳波中には常に混在するノイズ(定常雑音)が含まれる。これらのノイズを取り除くために定 Q 変換で得たスペクトログラムに対して、同サンプル全体の平均値を引き去る。これにより、定常的に発生する雑音を除去する。除去の計算式を式(5)に示す。ここで、 ch は電極、 k は周波数軸、 t は時間、 x はスペクトログラム、 $\overline{x(ch, k)}$ はスペクトログラムにおける全時間の平均である。

$$y(ch, k, t) = x(ch, k, t) - \overline{x(ch, k)} \quad (5)$$

3.3 想起区間の推定手法

評価実験では、想起反応が刺激からどのタイミングで発生するかを確認し、その区間を想起データとして用いた音声想起区間の推定を行う。

3.3.1 スペクトログラムの加算

脳波信号は微弱なため、本研究では被験者毎に 200 サンプルを加算し、音声想起反応を観察した。そして、予備実験で収集した発話音声のパワーの範囲を加算したスペクトログラムと比較することで、音声想起反応が現れる範囲を確定した。

3.3.2 音声想起区間推定

時系列に想起の有無を識別する。音声想起反応は時刻とともに複数の frame 上を変移しながら表出されると考え、識別器には 3 次元の畳み込みによる 3DCNN を採用した。

学習データは、想起と想起以外のスペクトログラムを 2 種類用意する。想起データは 3.3.1 で述べた加算スペクトログラムから得られる想起区間を用いる。また、想起以外のデータは想起区間の後方を採用する。想起反応が複数の frame 上を変移するという仮定から、7frame 窓、1frame シフトとして「周波数軸、7frame (窓長)、時間軸」の 3 次元のデータを作成し、3DCNN に入力する。3DCNN の構成を表 1 に示す。出力は想起か想起以外の 2 クラスである。このモデルをテストデータに適用し、時系列で想起か想起以外かの結果を得ることで、想起区間の推定を行う。また、20 分割交差検証で評価する。

表 1 3DCNN の構成

層	ユニット	出力形状	Filter-Size
入力	-	1*72*7*34	-
3次元畳み込み	16次元	16*70*6*25	3*2*10
活性化	ReLU	16*70*6*25	-
3次元畳み込み	32次元	32*68*6*16	3*1*10
活性化	ReLU	32*68*6*16	-
3次元 pooling	Maxpooling	32*13*6*5	5*1*3
dropout	0.20	32*13*6*5	-
Flatten	平坦化	12480	-
全結合	128次元	128	-
活性化	ReLU	128	-
全結合	2次元	2	-
活性化	Softmax	2	-

バッチサイズ：64、エポック数：30、最適化：Adam
損失関数：交差エントロピー

4 推定結果と考察

本実験の評価結果について述べる。実験試料は収録した日本語母音/a/として、各被験者 200 サンプルを使用する。

4.1 加算スペクトログラムの結果・考察

被験者 A のスペクトログラムを加算した結果を図 3 に示す。上図は予備実験で収録した音声発話を時間[sec]-パワーに変換したもの、下図は音声想起（発話無し）時のスペクトログラムである。横軸は刺激呈示終了時刻を 0 としている。この結果より、音声想起の反応は、音声発話に近いタイミングで開始/終了していることが確認できる。また、想起反応は急激に現れるものではなく、準備段階を経て徐々に生じているように見える。周波数帯域で比べると、特に 60~100Hz に強い成分が確認された。以上の想起反応はすべての被験者に同様の傾向が観察された。よって、この区間を音声想起の正解ラベルとし推定を行う。

4.2 音声想起区間推定の結果・考察

被験者毎に 20 分割交差検証で想起区間を推定した結果を図 4 に示す。図 4 は、横軸が時間、縦軸は被験者 5 名の 200 サンプル（計 1000 サンプル）に対する accuracy である。音声想起区間は赤枠の範囲である。先行研究[2]の手法（図 4 青線）は音声想起区間の accuracy が 0.2 まで低下しているが、提案手法（図 4 橙線）は安定して 0.8 の accuracy を得ている。この結果から、本手法は想起区間の推定に有効であると言える。しかし、提案手法の音声想起区間以外の accuracy に目を向けると、音声想起区間外を想起と推定してしまうことが先行研究をよりも多い。この要因は脳内のノイズによる影響であると考えられる。そのため、加算した 10 サンプルを本手法に適用することで、これらのノイズの低減を図った。結果は図 4 の緑線である。想起区間の精度は維持しつつも、想起区間外の精度が向上した。この結果より、本提案手法で最適な加算数を検討した。その結果を図 5 に示す。図 5 は各加算数における全被験者の accuracy を表しており、青線が全時間の平均 accuracy、橙線が想起区間の最大 accuracy を示す。また、横軸はサンプル加算数、縦軸が accuracy である。加算することで想起区間における精度が向上し、約 90% の accuracy を得た。また、10 回加算することで 1 サン

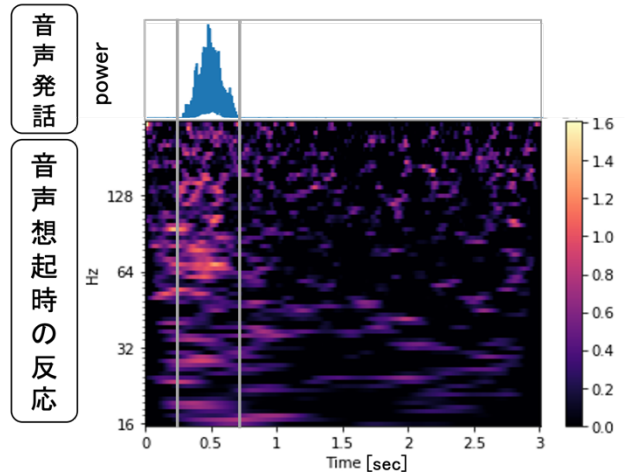


図 3 加算したスペクトログラム（被験者 A）

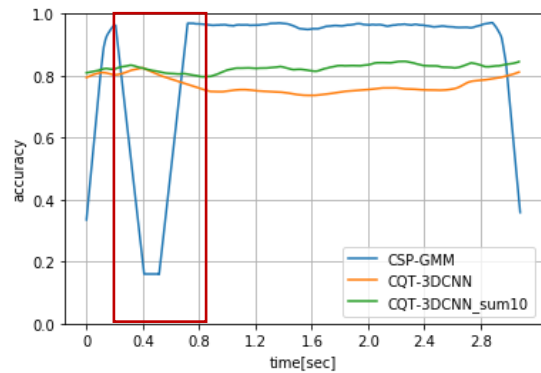


図 4 全サンプルの時間単位の識別率

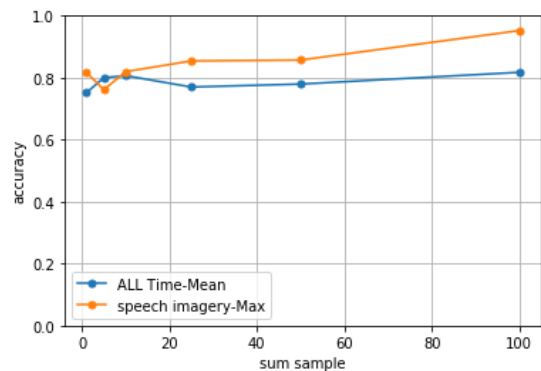


図 5 全時間の平均識別率と想起区間の最大識別率
プル時と比べ、全時間の平均識別率と想起区間の最大識別率が向上したことから、約 10 回加算するとノイズ低減が可能であると示された。

5 おわりに

本報告では、音声想起区間の検出を目的として、定 Q 変換によるスペクトログラムを用い、3DCNN で音声想起区間の推定を行った。その結果、データを加算することで約 90% の精度で想起推定に成功した。今後は被験者数とデータ数の拡充とともに、連続想起や刺激無し想起の区間推定を目指す。

参考文献

- [1] Tsuneo Nitta, Junsei Horikawa, Yurie Iribe, Ryo Taguchi, Kouichi Katsurada, Shuji Shinohara, Goh Kawai, “Linguistic representation of vowels in speech imagery EEG”, *Front. Hum. Neurosci.*, 18 May 2023 Sec. Brain-Computer Interfaces, Volume 17, 2023.
- [2] 細川晃樹, 山尾元陽, 入部百合絵, 田口亮, 新田恒雄, “共通空間パターンを用いた音声想起時 EEG の識別”, 情報処理学会第 83 回全国大会, 2021.
- [3] Judith C. Brown, “Calculation of a Constant Q Spectral Transform”, *J. Acoust. Soc. Am.* 89, 425 - 434. 1991.
- [4] Shuiwang Ji, Wei Xu, Ming Yang, Kai Yu, “3d convolutional neural networks for human action recognition”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol.35, 2013.
- [5] Anne Porbadnigk, Marek Wester, Jan-P. Callies, Tanja Schultz, “EEG-BASED SPEECH RECOGNITION Impact of Temporal Effects”, *BIOSIGNALS 2009 - Proceedings of the International Conference on Bio-inspired Systems and Signal Processing*, 2009.
- [6] Stephanie Martin, Peter Brunner, Inaki Iturrate, Jose del R. Millan, Gerwin Schalk, Robert T. Knight, Brian N. Pasley, “Word pair classification during imagined speech using direct brain recordings”, *Scientific Reports* volume 6, Article number: 25803, 2016.
- [7] Adeen Flinker, Anna Korzeniewska, Avgusta Y. Shestyuk, Piotr J. Franaszczuk, Nina F. Dronkers, Robert T. Knight, and Nathan E. Crone, “Redefining the role of Broca’s area in speech”, *PNAS* vol.112, no.9, pp.2871-2875, 2015.