

母音想起時における脳信号の周波数特性に基づいた想起区間検出

栗栖駿¹ 入部百合絵¹¹愛知県立大学 情報科学部

is211025@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研究が進められている。音声認識とは異なり、脳波による音声言語識別の研究では、想起区間が明確ではないため、広範囲に想起区間を対象としている。そのため、想起していない区間から言語情報を検出することになるため、識別精度の劣化が考えられる。そこで本稿では、想起言語の識別精度向上を目標に、5 母音を対象とした音声想起時脳波における想起区間検出について報告する。本報告では、定 Q 変換による時間周波数特性とコヒーレンス解析による位相同期を用いて想起と無想起の差異を測るとともに、それらの特徴をもとに 3DCNN により想起区間検出を行った。その結果、位相同期では、最大約 80% の精度で想起区間推定に成功したが、個人差が大きく生じていることが分かった。一方、時間周波数特性を用いた想起区間検出は、約 70% の精度を得た。

1 はじめに

近年、BCI(Brain Computer Interface)と呼ばれる、脳信号を用いてコンピュータを操作する技術が発展している。BCI に利用される脳信号として、発話せずに想起する音声想起(speech imagery)と呼ばれるものが注目を集めており、音声想起中の脳信号から語彙や音節、文章を推定する研究が進められている。音声想起中の脳波から精度よく音声言語を識別するためには、音声認識と同様にその対象区間の推定精度が重要である。しかし、未だに脳信号における想起と無想起の違いが明らかになっておらず、脳信号から想起区間を検出することは困難である。そのため、脳波から単語や音節を識別する先行研究の多くは、想起と仮定する区間を広範囲に設定して識別しているため、想起していない区間も識別対象としている可能性がある。

そこで、本研究では音声想起区間と無想起区間を識別することを目的とする。音声想起区間を明らかにすることで、想起時の言語特徴をより適切に分析

することが可能となり、単語や音節識別の高精度化に貢献できる可能性が高い。

先行研究では、音声想起と無想起の 2 クラス分類を行っているが、音声想起区間の検出までは行っていない[1]。一方、音声想起区間を検出している先行研究では、母音「/a/」の想起のみを対象としている[2]。本研究では 5 母音「/a/i/u/e/o/」を想起した際の想起区間検出を目標とする。

2 脳信号の収録

インフォームド・コンセントを得た右利き成人男性 8 名を対象に、日本語の母音「/a/i/u/e/o/」を想起した脳信号を収録した。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は 512Hz である。電極は拡張 10-20 法を使用し、言語に関する言語中枢周辺の左脳側に 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)、基準電極を左の耳朵、アース電極を AFz (図 1; 青丸)、g-HIAMP でインピーダンス測定を行うために必要となる Z 電極を FPz (図 1;

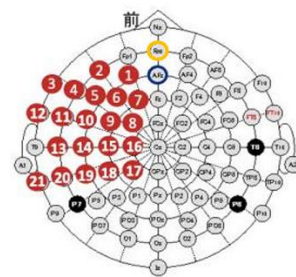


図 1 使用した電極配置(拡張 10-20 法)



図 2 収録プロトコル

黄丸)とした。被験者は一定の時間間隔で変化する画面表示に合わせて音声想起を行う。収録回数は各母音 20 セットであり、1 セットにつき 10 回連続で想起を行う。収録プロトコルを図 2 に示す。図 2 の + と表示されている時間帯は、画面に十字マーカーが表示されており、マーカーが消えた瞬間に想起を開始する。想起は一回行うように被験者に指示した。その際、母音の文字は思い浮かべず、音声発話を想像するように指示した。

3 提案手法

本報告で提案手法について、前処理、特徴量抽出、想起区間検出の順に詳述する。

3.1 前処理

収録したデータには、脳信号とは関係のないノイズが多く含まれているため、分析前にノイズ除去を行う。

3.1.1 直流成分除去

本研究で使用する脳波計は DC アンプであるため、収録された脳信号において直流成分のオフセットが混在している。直流成分は周波数解析に影響を及ぼす可能性が高いため、直流成分を除去する。本実験では、前後 50ms の区間毎に平均を計算し、元データから引き去ることで直流成分の除去を行う。計算式を(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 電源ノイズ除去

収録したデータには、電源由来のノイズが発生する。収録場所は愛知県であるため 60Hz とその倍音周波数に対しノッチフィルタを適用する。

3.1.3 電極間バイアス除去

脳信号は、基準電極と探査電極間の電位差を増幅したものである。そのため、電極の位置関係によって収録される電位は変化してしまう。そこで、基準電極の位置による影響をなくすため、測定された電極全体の電位の平均値を新たな基準として計算し、各電極から引き去った。

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。

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

3.1.4 21 電極加算

本研究では、21 電極で収録を行う。皮質脳波 (ECoG)の研究[3]で、音声想起反応のある電極にばらつきが生じていた。また、脳波計の電極にキャップ型を採用している影響があり、被験者の頭の大きさで電極の活性度に影響が出る。以上の事由により、本研究では 21 電極を加算する。

3.1.5 逆位相フィルタリング

21 電極加算により、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。そこで、想起とは関係のない強い反応を取り除くため、位相を用いたノイズキャンセリングを行う。ノイズキャンセリングでは、元の波形の位相スペクトルを反転させて振幅波形に戻すことで、元信号と反転した信号を生成し、元信号に加算することで特定の区間の音を打ち消せる。

3.2 特徴量抽出

3.2.1 脳波コヒーレンス

Lichtheim は、発話の流れを図式化したウェルニッケ・リヒトハイムの失語モデルを提唱した[4]。このモデルは、聴覚的に提示された言語を音声として出力するまでの経路を表しており、その中間地点でウェルニッケ野からブローカ野へ情報伝達が行われると示されている。この伝達は言語を発話する準備のために行われているものだと考えると、発話だけではなく想起の場合も同様に情報伝達が行われている可能性がある。想起と無想起間では、この情報伝達に相違があると仮定し、本研究では脳領域間の相関を分析する。

脳波コヒーレンスを用いた位相同期解析により、脳内部位の相関を求める。脳波コヒーレンスとは、2 つの脳部位間でそれぞれ得られた脳波に含まれる周波数成分毎の相関関係を示すものである。したがって、脳波コヒーレンスを調べることで、2 つの脳部位における脳信号の位相同期を求めることが可能になる。本研究では脳内部位の組み合わせとして、ブローカ野にあたる電極は 4ch とウェルニッケ野にあたる 19ch を選択した。一方、想起中は集中状態にあるとも考えられるため、ウェルニッケとブローカ野以外にも、集中時に活性化されている、前頭葉の周辺でも位相同期解析を実施した。前頭葉

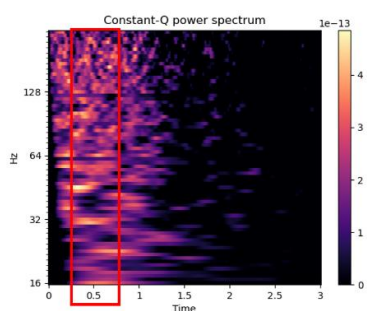


図3 加算スペクトログラム(被験者 A「/a/」)

の電極には 1ch と 6ch を選択した。

3.2.2 定 Q 変換 [5]

音声想起についての研究では、周波数帯域毎に想起反応との関係性を分析する研究が盛んに行われているが、想起反応と相関の高い帯域は未だ特定されていない。本研究では特定の周波数帯域に的を絞るのではなく、広範な周波数帯域を対象に音声想起反応を分析する。

スペクトログラムを用いる主な手法として、STFT(短時間フーリエ変換)が挙げられる。STFT では一定の窓長に基づき解析を行う。それにより、高周波数帯域に比べ、低周波数帯域では周波数分解能が下がってしまう。そこで、周波数分解能を一定にすることができる定 Q 変換を用いて、低い周波数帯域から高い周波数帯域 (16Hz~256Hz) について分析を行う。

定 Q 変換では、中心周波数毎に窓長を変えることが可能であるため、すべての周波数帯域で様な時間周波数分解能を得られる。

3.2.3 スペクトルサブトラクション

定 Q 変換で求めたスペクトログラムには、呼吸や心臓の鼓動、脳内で定常的に発され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)が含まれている。このノイズを取り除くために収録したデータに対して、同サンプル内の平均値を用いて定常雑音として引き去ることでノイズを除去する。

3.2.4 全サンプル加算

想起反応を増幅させるために、被験者内で母音毎にスペクトログラムを全加算する。スペクトログラムは定 Q 変換より得たものである。被験者 A の母音/a/の結果を図3に示す。なお、予備実験では想起区間の参考とするために音声発話も収録した、赤枠

は同被験者の発話開始から終了までの発話区間を示している。図から、発話区間とスペクトログラムの成分の増減が重なっており、定 Q 変換では想起と考えられる区間を捉えることができていると考えられる。音声想起反応が音声発話に近いタイミングで発生していることが分かる。

3.3 想起区間検出手法

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 上で変移しながら表出されると考えられる。そのため、音声想起反応の時間変化を考慮し、識別器に 3DCNN を用いる。学習データは、想起と無想起の 2 種類を用意する。

予備実験で得た音声発話では、発話開始の合図から発話まで間が開いていた。また、3.2.4 から音声想起反応が音声発話に近いタイミングで発生していると考えられる。よって、学習と評価に用いる想起の区間は、予備実験での音声発話区間の時刻とする。また、無想起区間はスペクトログラム内で成分の見られない、想起区間の後方とする。コヒーレンスと定 Q 変換より得たスペクトログラムをそれぞれ、7frame 窓、1frame シフトして、「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。データは、1 被験者 1 母音毎に 200 個あり、学習データを 190 個、テストデータを 10 個とする。出力は想起か無想起の 2 クラスである。このモデルをテストデータに適用し、時系列で想起か無想起かの結果を得ることで、想起区間の検出を行う。また、20 分割交差検証で評価する。

4 検出結果と考察

4.1 位相同期を用いた区間検出

コヒーレンス解析から得たスペクトログラムを用いて想起区間を検出した。周波数帯域は、高次認知機能に係る 50-100Hz と集中に係る $\beta \cdot \theta$ 波を対象とした。3DCNN の構成を表 1 に示す。バッチサイズ 64、エポック数 30 とし、最適化は Adam 損失関数による交差エントロピーである。

4ch-19ch における位相同期を用いた想起区間検出の結果を図 4、1ch-6ch の結果を図 5 に示す。50-100Hz では良好な結果が得られなかったため、紙幅の関係上、 $\beta \cdot \theta$ 波を用いた結果を示した。図は時間毎に想起と判定される割合である (再現率)。そのため、

表 1 3DCNN の構成 (位相同期)

層	ユニット	出力形状	フィルタサイズ
入力層	-	42*7*34(*1)	-
3次元畳み込み層	16 次元	38*5*30*16	5*3*5
活性化層	ReLU	38*5*30*16	-
3次元畳み込み層	32 次元	34*3*26*32	5*3*5
活性化層	ReLU	34*3*26*32	-
3次元 pooling 層	Maxpooling	6*3*8*32	5*1*5
dropout 層	0.40	6*3*8*32	-
Flatten 層	平坦化	4608	-
全結合層	128 次元	64	-
活性化層	ReLU	64	-
全結合層	2 次元	2	-
活性化層	Softmax	2	-

想起と判定される場合は 1 に近づき、無想起の場合は 0 に近づく。赤枠は被験者の発話区間を参考にしたものである。青枠は、想起以外のデータとして用いた区間である。

被験者の一部は、特定の母音に対する再現率が 80% を超える結果が得られた。一方、被験者 D のように想起区間の精度が低い例も観測され、個人差が生じる結果となった。

4.2 時間周波数特性を用いた区間検出

次に時間周波数特性を用いて区間推定を行った。3DCNN のモデル構造は位相同期時と同様であり、1,2 層目の 3 次元畳み込み層のフィルタサイズを (3,3,3) に変更した。バッチサイズ、エポック数、最適化は位相同期と同様である。

被験者の区間推定結果を図 6 に示す。赤枠は、全被験者の発話区間である。

紙幅の関係上、時間周波数特性については被験者毎の結果を割愛しているが、位相同期では結果に個人差が生じていたのに対し、時間周波数特性を用いた場合は、全被験者が発話区間内の値が 1 に近づき、

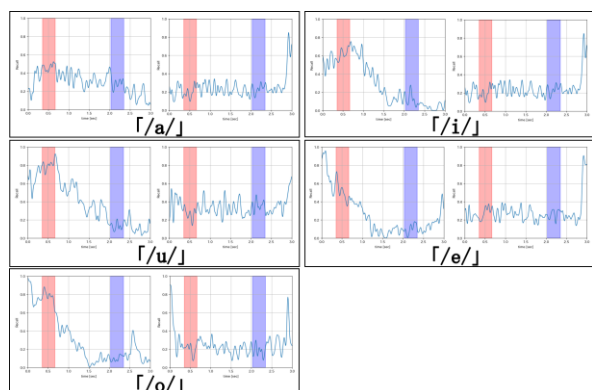


図 4 4ch-19ch 間の位相同期を用いた区間検出結果 (「左: 被験者 A」 「右: 被験者 D」)

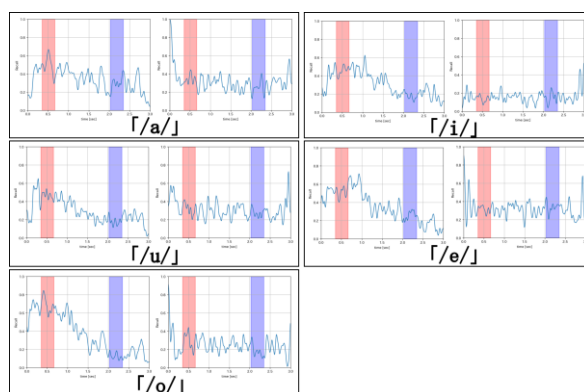


図 5: 1ch-6ch 間の位相同期を用いた区間検出結果 (「左: 被験者 A」 「右: 被験者 D」)

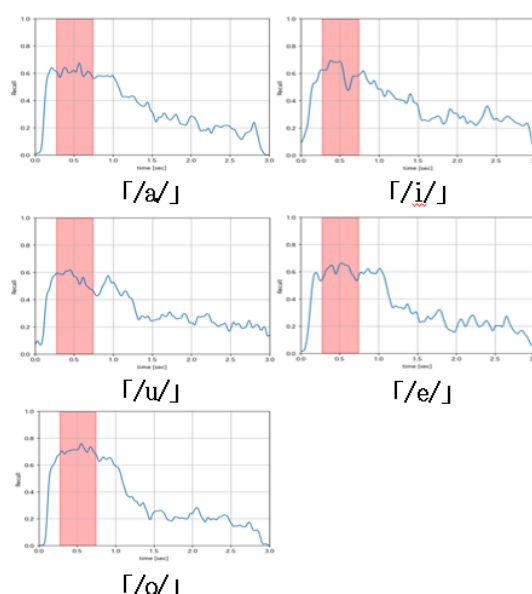


図 6 全被験者の時間周波数特性を用いた

想起と判定される割合が高く、信号後方にかけては 0 に近づき無想起と判定される傾向を確認できた。また、再現率が母音によっては 7 割を超えた。そのため、位相同期と比較すると、定 Q 変換による周波数特性のほうが安定して想起区間を検出できる結果となった。

5 おわりに

脳内における想起と無想起の状態の特性を取得するために、時間周波数特性と位相同期に着目したが、本結果より時間周波数特性の一つである定 Q 変換が有効であることが分かった。

参考文献

- [1] M. AlSaleh, R. Moore, H. Christensen and M. Arvaneh, "Discriminating Between Imagined Speech and Non- Speech Tasks Using EEG", 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Honolulu, HI, USA, pp. 1952-1955, 2018.
- [2] 鈴木大祐, "音声想起時の脳波における想起区間の推定", 言語処理学会 第 30 回年次大会 発表論文集, pp.1985-1989, 2024.
- [3] 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.
- [4] 廣谷定男, "聞くと話すの脳科学", コロナ社, 2017.
- [5] Judith C. Brown and Miller S. Puckette, " An efficient algorithm for the calculation of a constant Q transform", J. Acoust. Soc. Am., 92(5), pp.2698-2701, 1992.