

# 調音運動前の言語中枢間の位相同期を用いた母音認識

長瀬南帆<sup>1</sup> 入部百合絵<sup>1</sup>

<sup>1</sup>愛知県立大学 大学院情報科学研究科

im232007@cis.aichi-pu.ac.jp iribe@ist.aichi-pu.ac.jp

## 概要

近年、発話したときの脳波を言語に変換する Brain to Text の研究が進められている。このシステムには、舌や口唇などの調音器官の障害により明瞭な発音が困難である人々への意思伝達支援が期待されている。しかし、脳波は信号雑音比が低い上、調音器官の運動、すなわち調音運動による筋電図信号が重畠されるなど、雑音の重畠が課題である。そこで本研究は、雑音の少ない調音運動前の脳波から日本語 5 母音を認識することを目的とする。特に、言語を音声として出力するための準備を担うブローカ野とウェルニッケ野の 2 つの言語中枢に着目する。調音運動前の言語中枢間の位相同期を用いることで、日本語 5 母音の認識精度が 69.6% を達成した。

## 1 はじめに

Brain to Text には、実際に言葉を話す発話を対象とした研究と、相手に話しかけるイメージで頭に思い浮かべる想起を対象とした研究がある。本研究は、加齢や病気が原因で「声は出るが発音が不明瞭」のように、正確な発話が困難である人々への意思伝達支援を念頭に置き、発話を対象とした。

Denise ら[1]は、調音運動開始から 1500[ms]間の脳波を言語に変換することを試みた。パワースペクトル密度 (PSD) を特徴量として、サポート・ベクトル・マシン (SVM) で 5 種類のスペイン語単語を認識した結果、正解率は約 41.8% であった。その精度は十分ではなく、改善の余地がある。

他の研究も含めた先行研究の多くは、精度に悪影響を及ぼす要因として雑音の重畠が挙げられる。脳波は微小な信号であり、信号雑音比が低い。さらに、発話には舌や口唇などの調音器官の運動、すなわち調音運動を伴うため、調音運動による筋電図信号が発生する可能性がある。つまり、発話をを行うことで脳波に含まれる雑音が増加することになる。これにより、脳波に含まれる言語に関わる特徴を捉え

られないため、認識精度が停滞していると考えられる。

そこで、調音運動前のブローカ野とウェルニッケ野の 2 つの言語中枢に着目した。調音運動前は言語を音声として出力するための準備段階であるため、筋電図信号は発生しづらい。つまり、調音運動前は調音運動の開始以降に比べて雑音の重畠が抑制されると考えられる。さらに、言語中枢であるブローカ野とウェルニッケ野は、調音運動前に活動することが明らかにされている。Flinker ら[2]は、皮質脳波と呼ばれる侵襲型の脳信号を解析し、調音運動の開始に向けてウェルニッケ野、ブローカ野の順に活性化することを示した。さらに Lichtheim[3]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。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が伝達されることを示している。以上のことから、調音運動前の言語中枢間の連携区間に言語に関わる特徴が含まれるのではないかと考えた。

したがって本稿では、脳領域間の相関を調べる位相同期解析で調音運動前の言語中枢間の連携を観察し、それらを用いて日本語 5 母音の認識を試みた結果を述べる。また、先行研究[1]の手法と比較した結果も併せて報告する。

## 2 調音運動前の脳波と筋電図の計測

インフォームド・コンセントを得た右利き成人男性 2 名 (A, B) の協力を仰ぎ、日本語 5 母音/a/, /i/, /u/, /e/, /o/ の脳波と筋電図を計測した。計測機器はミユキ技研のポリメイト V AP5148 を使用した。サンプリング周波数は 1000[Hz] で、60[Hz] および倍音周波数の電源ノイズを除去するノッチフィルタが搭載されている。電極配置は拡張 10-20 法に従い、言語中枢が存在する左脳側に 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,

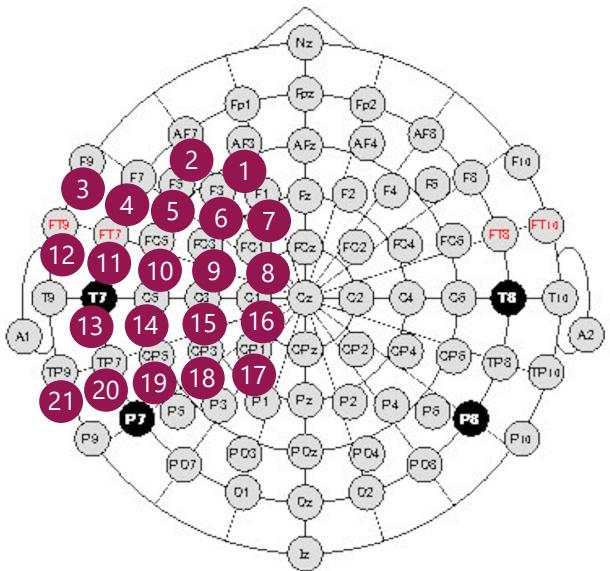


図 1 左脳側 21 電極の配置

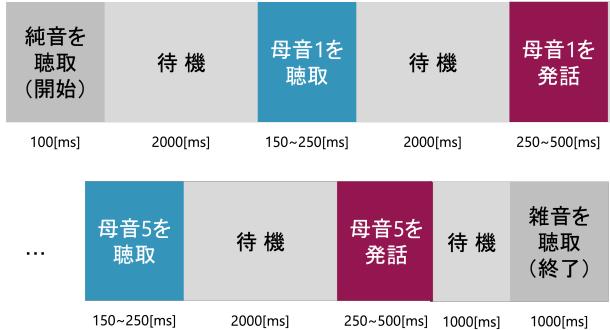


図 2 実験プロトコル

16; C1, 17; CP1, 18; CP3, 19; CP5, 20; TP7, 21; TP9) を図 1 のとおり配置した。また、基準電極を左耳朶、アース電極を前額部、筋電図用の電極を左右の口角下制筋に装着した。

図 2 の実験プロトコルに沿って日本語 5 母音の発話を行った。実験は予め用意した母音音声を聴取した後、同じ母音を発話するという流れで進行した。5 母音の発話を各 1 回終えたところで実験を終了する。これを 1 セットとして 200 セット実施し、被験者 1 名につき 1 母音を 200 サンプルずつ採取した。

### 3 前処理

脳波には体動や呼吸などの生体現象、計測機器や環境に起因する雑音が含まれている。これらは解析の妨げとなるため、前処理として雑音除去を行った。

#### 3.1 電極再基準化

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であるため、電極の位置関係による偏りが

生じる。そこで、全電極の平均値を各測定電極から減算する電極再基準化を行った。計算式を式 (1) に示す。 $ch$ は電極、 $t$ は時間、 $x(ch, t)$ は各電極の時間波形、 $\bar{x}(t)$ は全電極を平均した時間波形である。

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

#### 3.2 直流成分除去

計測された電位には脳波計由来の直流成分が含まれている。そこで、各フレームの前後 50[ms]間の平均値を減算することで直流成分を除去した。計算式を式 (2) に示す。 $t$ は時間、 $x(t)$ は時間波形、 $\bar{x}$ は  $t$  の前後 50[ms]間の平均値とする。

$$y(t) = x(t) - \bar{x} \quad (2)$$

#### 3.3 スペクトルサブトラクション

筋電図信号による影響を除去するために、スペクトルサブトラクションを行った。筋電図に 10-400[Hz]のバンドパスフィルタを適用し、窓長を 64[ms]、シフト長を 16[ms]として、脳波の振幅スペクトルから筋電図の振幅スペクトルを減算した。

### 4 位相同期解析

調音運動前の言語中枢間に生じる連携の有無と強度を観察するために、位相同期解析を行った。言語中枢のブローカ野 (ch.4) とウェルニッケ野 (ch.19) に加えて、運動に携わる一次運動野 (ch.9, ch.10) も対象とした。

#### 4.1 位相同期度 (PLV) [4]

位相同期度 (PLV) は、2 つの脳領域間の位相の相関の時間変化を表す指標である。PLV は 0 から 1 の範囲で値をとり、相関が高いほど 1 に近づく。計算式を式 (3) に示す。 $N$  は試行数、 $t$  は時間、 $\phi_1(t, n)$  と  $\phi_2(t, n)$  は  $n$  試行目の位相である。本研究では、ヒルベルト変換で抽出した瞬時位相を使用した。

$$PLV(t) = \left| \frac{1}{N} \sum_{n=1}^N e^{j(\phi_1(t, n) - \phi_2(t, n))} \right| \quad (3)$$

#### 4.2 PLV の結果と考察

被験者 A の PLV の結果を図 3 に示す。ブローカ野 (ch.4) とウェルニッケ野 (ch.19)、ブローカ野と一次運動野 (ch.9, ch.10) の PLV を、上段から/a/, /i/, /u/, /e/, /o/ の順で記載した。時間軸は、調音運動前 1000[ms] と調音運動の開始以降 500[ms] の範囲

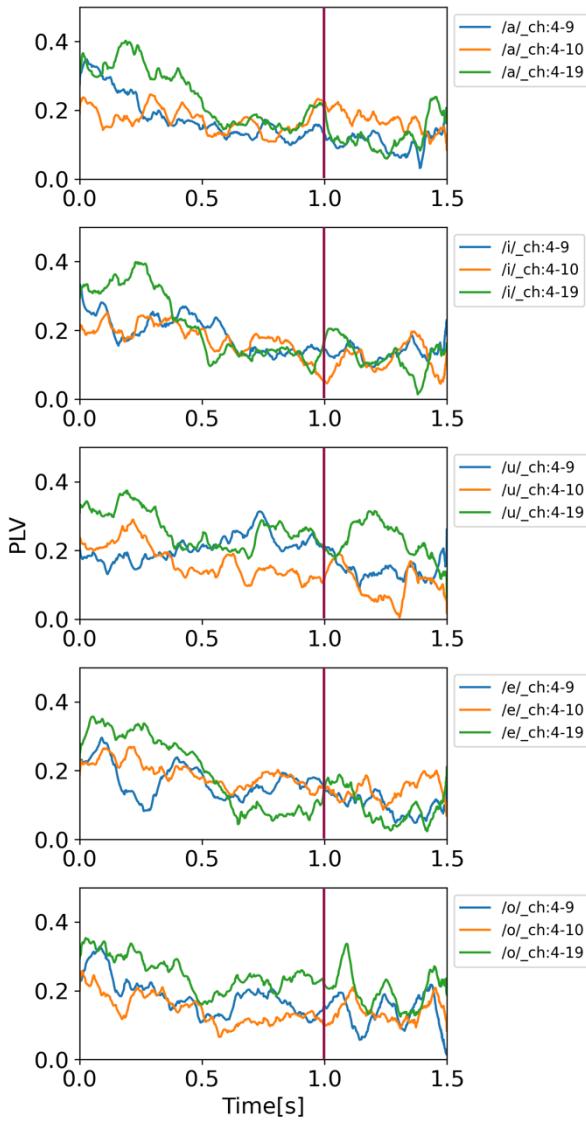


図 3 被験者 A の PLV

を示しており、各グラフの 1.0[s] の地点が調音運動の始端を表している。調音運動前に注目すると、どの母音においてもプローカ野 (ch.4) とウェルニッケ野 (ch.19) の PLV が高いことは明らかである。さらに、音声が outputされる調音運動開始以降より、調音運動前の PLV が高い。以上のことから、調音運動前の言語中枢間に言語を音声として出力する準備を行うための連携が生じていると考察した。

## 5 位相同期を用いた母音認識

位相同期解析にて、調音運動前の言語中枢間に連携が生じることが示された。そこで、言語中枢間の連携区間において言語に関わる特徴が含まれていると仮定し、位相同期を用いた母音認識を試みた。

### 5.1 特徴量抽出

特徴量として、プローカ野とウェルニッケ野の PLV、PLV の 1 次微分、PLV の 2 次微分、プローカ野の瞬時振幅、ウェルニッケ野の瞬時振幅を抽出した。PLV の導出は 4.1 節の式 (3) のとおりである。また、瞬時振幅はヒルベルト変換で抽出した。

特徴量を抽出する区間は調音運動前 1000[ms] である。ただし、次元圧縮のために 250[ms] ごとに平均し、4 次元に抑えた。また、母音認識時のサンプル数を確保するために、試行平均の回数を 25 回にした。

さらに、先行研究[1]の手法との比較を行うために、パワースペクトル密度 (PSD) を抽出した。先行研究[1]に倣い、PSD を 1-4[Hz], 4-7[Hz], 8-13[Hz], 12-15[Hz], 15-20[Hz], 18-30[Hz] の計 6 つの周波数帯域に分割し、帯域ごとに平均値を計算した。その他の条件は本研究と同様で、抽出する区間は調音運動前 1000[ms]、対象の脳部位はプローカ野とウェルニッケ野である。

### 5.2 調音運動前からの母音認識手法

被験者ごとに母音認識を行った。実験試料は、各被験者 1 クラス 100 サンプル、計 500 サンプルである。各被験者の 1 母音につき 200 試行あるサンプルから、ランダムに選択された 25 試行を平均したサンプルを 100 個生成した。平均されるサンプルの重複を許することで、サンプル数を増加させた。識別器にはサポート・ベクトル・マシン (SVM) を使用し、10 分割交差検証でモデルを評価した。評価指標は正解率と F 値である。

### 5.3 母音認識の結果と考察

母音認識の結果を表 1 から表 4 に示す。表 1 と表 3 は被験者 A、表 2 と表 4 は被験者 B の結果である。先行研究[1]の手法を用いた結果は、先行研究[1]よりも認識精度が大幅に上昇していることがわかる。本研究では脳波の信号雑音比を改善するために加算を行った。それが、母音認識の妨げとなる雑音を減少させたことで、認識精度が向上したと考えられる。

被験者 2 名ともに、先行研究[1]の手法に比べて、提案手法の認識精度の方が約 14% 高い正解率であった。さらに、先行研究[1]の手法に PLV を加えることで、認識精度が約 20% 上昇した。以上のことから、PLV を組み合わせることで母音間の差異が強調され、認識精度が向上したと考えられる。

**表 1 被験者 A の母音認識の結果**

手法	正解率[%]	F 値[%]
提案手法		
• PLV		
• PLV の 1 次微分	76.8	75.9
• PLV の 2 次微分		
• 瞬時振幅		
先行研究[1]	63.2	57.6
• PSD		
先行研究[1]+提案手法		
• PSD		
• PLV		
• PLV の 1 次微分	81.6	81.2
• PLV の 2 次微分		
• 瞬時振幅		

**表 2 被験者 B の母音認識の結果**

手法	正解率[%]	F 値[%]
提案手法		
• PLV		
• PLV の 1 次微分	62	61.8
• PLV の 2 次微分		
• 瞬時振幅		
先行研究[1]	48	41.9
• PSD		
先行研究[1]+提案手法		
• PSD		
• PLV		
• PLV の 1 次微分	70	69.5
• PLV の 2 次微分		
• 瞬時振幅		

**表 3 先行研究[1]+提案手法による被験者 A の認識結果の混同行列**

		予測値				
		/a/	/i/	/u/	/e/	/o/
真値	/a/	42	1	0	5	2
	/i/	3	46	0	1	0
	/u/	0	0	40	0	10
	/e/	10	6	0	34	0
	/o/	0	0	8	0	42

**表 4 先行研究[1]+提案手法による被験者 B の認識結果の混同行列**

		予測値				
		/a/	/i/	/u/	/e/	/o/
真値	/a/	41	0	3	3	3
	/i/	2	39	3	4	2
	/u/	2	4	33	6	5
	/e/	7	6	4	26	7
	/o/	3	3	3	5	36

調音運動は音源に言語としての音響的特性を付与する際に生じる現象であるため、言語に関わる特徴が最も濃厚に反映されている区間は調音運動の開始以降であると推測できる。しかし調音運動前の脳波を用いる提案手法は、調音運動の開始以降の脳波を使用した先行研究[1]の認識精度を上回った。本研究は、人間が発話する意思を持つ前から言語関連の脳活動が生じることを示すとともに、それらが母音ごとに異なる性質を有している可能性を見出した。

## 6 おわりに

本稿では、位相同期解析で調音運動前の言語中枢間の連携を観察し、それらを用いて日本語 5 母音の認識を試みた結果を報告した。位相同期解析では、調音運動前の脳波において言語中枢間に相関が生じることが示された。さらに、調音運動前の方が調音運動の開始以降に比べて相関が高いことから、調音運動前の言語中枢間には調音運動の準備を行うために連携が生じていると考察した。位相同期解析の結果を踏まえて、調音運動前の言語中枢間の位相同期を用いた母音認識を試みたところ、被験者平均で正解率 69.6% の認識精度を獲得し、先行研究[1]の手法を上回った。

今後は被験者数とデータの拡充とともに、コーリンスなどの他の位相同期解析に取り組み、さらなる精度向上を目指す。

## 参考文献

- Denise, Alonso-Vazquez; Omar, Mendoza-Montoya; Caraza, Ricardo; Martinez, Hector R.; Antelis, Javier M.. EEG-Based Classification of Spoken Words Using Machine Learning Approaches. Computation. 2023, vol.11, no.11, p.225-244.

2. Flinker, Adeen; Korzeniewska, Anna; Shestyuk, Avgusta Y.; Franaszczuk, Piotr J.; Dronkers, Nina F.; Knight, Robert T.; Crone, Nathan E. Redefining the role of Broca's area in speech. *Proceedings of the National Academy of Sciences*. 2015, vol.112, no.9, p.2871-2875.
3. Lichtheim, Ludwig. On Aphasia. *Brain*. 1885, vol.7, no.4, p.433-484.
4. Lachaux, J. P.; Rodriguez, E.; Martinerie, J.; Varela, F. J. Measuring phase synchrony in brain signals. *Human Brain Mapping*. 1999, vol.8, no.4, p.194-208.