

脳波中の母音想起信号に対する認識精度向上 Improvement of recognition accuracy for imagined vowel in EEGs

鈴木 裕毅^{†1} 篠原 修二^{†1} 高橋 達二^{†1} 田口 亮^{†2} 入部 百合絵^{†3} 新田 恒雄^{†4}
Hiroki Suzuki Shuji Shinohara Tatsuji Takahashi Ryo Taguchi Yurie Iribe Tsuneo Nitta

1. はじめに

音声言語に基づく BCI (Brain Computer Interface) の研究開発を行っている。発話時 EEG (electroencephalogram) 信号は、音声器官への指令 (motor command) を利用できるが、想起時 EEG 信号は言語表象 (linguistic representation) 自体を同定する必要がある。本報告では、我々がこれまで行ってきた脳波中言語表象に関する研究のうち、想起短音節中の母音認識と認識精度向上を中心に報告する。

2. 短音節音声想起脳波からの母音認識

音声言語は、中側頭回 (MTG) で言語としてのプランニングが行われた後、MTG 後部から上側頭回 (STG) にかけて音韻コードが選択され、下前頭回 (IFG: ブローカ野) において音節化 (syllabication) がなされる。音声発話ではこの後、中心前回 (PG) で調音動作指令が出される [1]。EEG 信号については、発話時を含む筋動作を伴う場合、空間情報 (spatial information) と時間情報 (temporal information) の抽出が重要とされる。一方考えている際、すなわち音声言語想起時は、上に述べたように音韻コードがブローカ野へ送られることを考えると、言語表象 (linguistic representation) は時間情報、具体的には周波数スペクトラム情報として EEG 信号に畳込まれるとするのが自然と考える。

2.1 脳波の線形予測分析

図 1 は脳波 $X(k)$ が、情報源の白色雑音 $W(k)$ と言語情報 $L(k)$; k : 周波数との畳込みにより生成されるモデルを示している。21 電極から成る脳センサーの各 EEG 信号 $X(k)$ 即ち言語情報 $L(k)$ は、線形予測分析 (Linear predictive analysis; LPA) により、(1), (2) 式から求めることができる。

$$L(k) = 1/\mathcal{F}\{a_0, a_1, a_2, \dots, a_8, 0, \dots, 0\} \quad a_0 = 1$$

$$= 1/\{\text{Re}X(k) - j \text{Im}X(k)\} \quad (1)$$

$$|L(k)| = |X(k)| = 1/\{\text{Re}^2X(k) + \text{Im}^2X(k)\}^{1/2} \quad (2)$$

ここで、 $\mathcal{F}\{\}$ は離散フーリエ変換 (DFT) を示す。 $\{a_p\}$ は線形予測係数で EEG 波形の自己相関係数を元に、Durbin の再帰アルゴリズムから求められる [2]。式 (1) は全極モデルと呼ばれる。

2.2 音声想起スペクトラムと線スペクトラム

図 2 に、数字 “6” /roku/ と “0” /zero/ を音声想起した際の線スペクトラム (line spectrum) を示す。線スペクトラムは LPA の振幅スペクトラム包絡から、1 次微分 (極大値抽出) と 2 次微分 (変曲点抽出) を通して求めている。

想起音声は、発音器官動作の拘束を受けないため、図のように離散的短音節 CV (C: 子音, V: 母音) の塊 (chunk) として観察される。

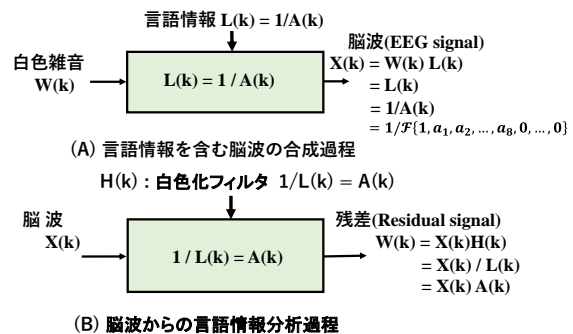


図 1 線形予測分析に基づく脳波分析

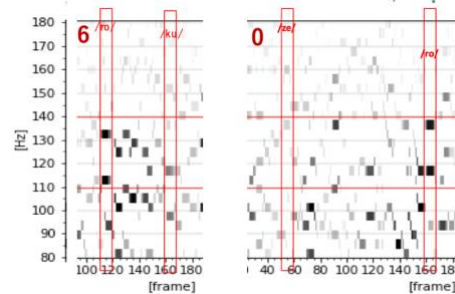


図 2 数字音声想起パターンの例 (線スペクトラム)

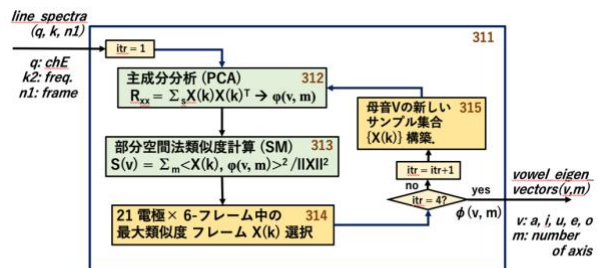


図 3 21電極×6frameからの母音線スペクトラム探索

3. 想起母音中の言語表象抽出

3.1 想起音声の中の短音節ラベリング

図 2 の濃淡パターンは、21 電極 (g.tec 社脳波計 g.HIAMP 使用) 毎に EEG 信号を周波数領域でノイズ除去した後、波形に戻し、線形予測分析結果の線スペクトラム $X(k)$ を電極間で pooling 処理したものである。図に赤枠で示した 9-frame (1-frame は 8msec) の区間は、目視による短音節ラベリングの結果を示す。個々の母音はこれらの区間に対応する 21 電極から抽出することになる。

3.2 想起短音節中の母音抽出

図 3 に離散短音節中の母音フレームを抽出し、固有ベクトルセット $\phi(V, m)$; $V=a, I, u, e, o, m$ =軸番号を獲得する探索手順を示した。これまで CV サンプルの観察から、母音区間は 9-frame

^{†1} 東京電機大学 Tokyo Denki University.

^{†2} 名古屋工業大学 Nagoya Institute of Technology.

^{†3} 愛知県立大学 Aichi Prefectural University

^{†4} 豊橋技術科学大学 Toyohashi University of Technology.

中の後半 6-frame, 子音区間は前半 6-frame (母音と子音は区間が overlap している) に存在することが明らかになっている。母音フレーム探索では, 特徴を視認可能な主成分分析(PCA) の固有ベクトル $\phi(V, m)$ 計算と, 21 電極 \times 6 フレーム中の適切なスペクトラムデータ $\{X(k)\}$ を $\phi(V, m)$ から評価できる部分空間法 (Subspace Method; SM) を繰り返し適用する。iterative search により $\phi(V, m)$ 再設計をルーチン化できる[2]。

3.3 母音言語表象

母音は {Ca, Ci, Cu, Ce, Co; C = #, k, s, t, n, h, m, r, g, z} と後続母音と同じ短音節から, $\phi(V, m)$ を用いて自動的に切り出し, 3.2 の母音スペクトラム探索を行い, 固有ベクトルを更新することができる。次に, 母音言語表象を母音固有空間から考えると, 母音個々の母音を代表する参照ベクトル $\text{ref}(V)$ は, 固有値 $\lambda(V, m)$ および固有ベクトル $\phi(V, m)$ から次式から表現できる。

$$\text{ref}(V) = [\sum_{m=1}^8 \{ \lambda(V, m) / \lambda(V, 1) \} \phi(V, m)^2]^{1/2} \quad (3)$$

$\text{ref}(V)$ は $\lambda(V, m) / \lambda(V, 1)$ を重みとする累積スペクトラムで, $\lambda(V, m)$ の値は $\phi(V, m)$ が $\text{ref}(V)$ へ寄与する割合を表す。図 4 に 5 母音の参照ベクトルを示した。次に, 音声信号の母音ホルマント ($F1, F2$) を念頭に, 想起母音の参照ベクトルから母音毎に二つのスペクトラムピーク ($P1, P2$) を抽出し, 図 5 にチャートとして示す。音声の ($F1, F2$) は, 声道の共鳴周波数を反映するため, 5 母音が 5 角形を構成し, 男, 女, 子供で大きく異なる。これに対して, $P1$ - $P2$ チャートは直線上に乗り, 男女差は見られない。

4. 想起短音節中の被検者独立母音認識

4.1 実験試料

被検者はインフォームド・コンセントを得た右利きの男声 4 名と女性 1 名の計 5 名である。実験は評価に 1 名, 残り 4 名を学習に用いる。

4.2 被検者独立母音認識実験

線スペクトラム (LS) 系列を CNN に入力し, 5 母音を分類する実験を行った。CNN は画像データ処理で優れた結果を得ているが, EEG 信号の場合も母音類似度の高い線スペクトラム LS を連続 6-frame 抽出して入力することにより, CNN の特徴を活かすことができる。表 1 に結果を示した。

4.3 調音特徴空間への写像による母音認識

脳波の母音線スペクトラムは, 図 4 から分かるように周波数成分間で相関を持つ。そこで成分間ができるだけ分離する本質的な特徴量として, 表 2 に示す調音特徴 (articulatory feature; AF) を導入する。母音の AF は, 開口の度合 (狭・半狭・広) と舌の位置 (前・中舌・後) を表現する。今回は線スペクトラム LS から調音特徴 AF への写像関数として CNN を用いた結果を表 1 (CNN-AF) に示した。CNN 単体と比較し, 特徴抽出器である CNN-AF と分類器としての MLN を用いた母音認識は 2.5% のゲインを得た。

調音特徴は音声認識でも利用され, 音声認識, 音声合成の双方に調音特徴を用いた, one-model 音声認識・合成システムが提案され, 高い性能を得られている [3]。

参考文献

[1] Levelt W., "Speaking: From Intention to Articulation" (ACL-MIT Series in Natural Language Processing) (1993).

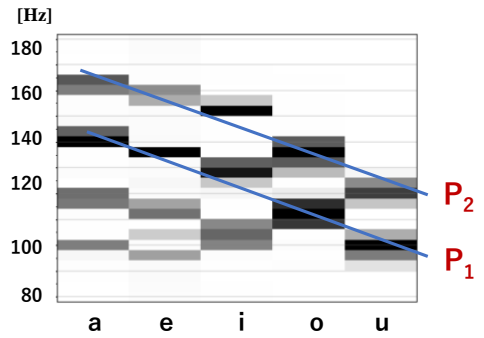


図 4 5 母音の参照ベクトル $\text{ref}(v)$; $v = a, e, i, o, u$

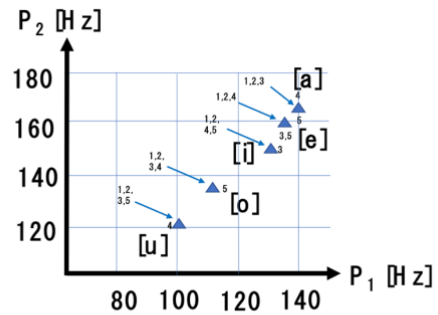


図 5 母音の $P1$ - $P2$ チャート (被検者 5 名)
1~4: male, 5 female $\Delta f = 3.9\text{Hz}$

表 1 5 母音認識結果

被検者	CNN	CNN-MLN(AF)
M1	69.3	74.0
M2	72.6	71.2
M3	74.4	77.5
M4	66.3	67.7
F1	69.4	71.2
全体精度	70.2	72.7

表 2 母音調音特徴表

母音	i	e	a	o	u
狭母音	1	0	0	0	1
半狭母音	0	1	0	1	0
広母音	0	0	1	0	0
前舌母音	1	1	0	0	0
中舌母音	0	0	1	0	0
後舌母音	0	0	0	1	1

5. おわりに

音声想起時の EEG 信号から母音を認識する方法を述べた。線スペクトラム系列の CNN による母音分類では, 調音特徴空間へ写像後, MLN により母音分類することで性能が改善することを報告した。

[2] Nitta T., et al., "Linguistic representation of vowels in speech imagery EEG", Frontiers in Human Neuroscience (May, 2023).

[3] Nitta, T., et al., "One-model speech recognition and synthesis based on articulatory movement HMMs", INTERSPEECH, pp.2970-2973 (2010).