

脳磁界データからの想起音声の判別に係る特徴量の推定

- ウェーブレット変換と SVM による解析 - *

☆宇澤志保美 (神戸大/産総研), 滝口哲也, 有木康雄 (神戸大), 中川誠司 (千葉大/産総研)

1 はじめに

近年, 脳活動を機械制御や意思伝達手段として活用する試みが生んで行われている。例えば, 注意を向けた刺激に対してのみ出現する誘発反応 (P300) によって, 間接的に意思伝達を行うブレイン・コンピュータ・インターフェース (BCI) の開発例は多い。ただし, この方式 (P300 スペラー型) [1] ではあらかじめ選択肢として用意されたもの以外の行為や事象を伝えることはできない。より汎用性のある BCI の開発のためには, ヒトの脳活動から認知や意思, 情動そのものを解読できることが望ましい。

近年, 音声想起時の脳活動の推定が試みられている。例えば, シラブルを想起させた時の fMRI 計測においては, 音声の想起を訓練した被験者のみに対する両側上側頭回の活動が報告されている [2]。また, 物体の打撃音を想起した時の MEG 計測においては, 右半球優位の緩やかな反応が聴覚野周辺に報告されている [3]。しかし, 活動の空間特性にはバラツキが大きい, 定性的なタスクについての事前トレーニングが反応に影響するなど, 高い再現性を持った反応波形が得られているとは言い難い。

我々はこれまで, 脳磁界計測を用いて, 音声の聴取時と比較した音声想起時の時空間特性の検証を行ってきた。音声聴取時には呈示音声の振幅エンベロープと脳磁界反応の間にある程度の相関があるものの, 音声想起時はそうした相関が得られなかった [4]。一方, 音声想起時にも聴覚野周辺の緩やかな活動は認められており, その活動の中には想起音声に起因する特徴量が含まれていると考えるべきであろう。つまり, 振幅エンベロープ以外のパラメータにも着目した, 発展的な解析を行う必要があると思われる。他方, 運動の想起に係る脳活動計測の報告は比較的多い [5]。運動想起時には自発脳活動の α (8-13 Hz), β (13-25Hz) 帯域が変動する [6] という報告がなされているが, 発話も一種の運動であるため, 音

声の想起時にも同様に周波数変化が捉えられる可能性がある。

音声想起時の脳反応は振幅が小さく, かつ緩やか (時間的同期性が悪い) であり, 時間波形上での解析や発生源推定のみから再現性のある特徴量を見いだすのは容易ではない。一方, 機械学習を用いることで, 微細な特徴量の変化を捉えることが可能になると考えられる。本稿では, 上記の問題点を踏まえて, 周波数特徴量 (連続ウェーブレット解析) と機械学習 (Support Vector Machine: SVM) による音声想起時の脳活動の特性評価を行った。SVM による学習モデルの生成に用いた特徴量次元の重みから, 想起音声の判別に有効なデータの潜時およびチャンネルを推定した。

2 計測方法

2.1 刺激呈示

ひらがなの単語刺激 (“あまぐも”, “いべんと”, もしくは “うらない”) を用いた。3 単語のうち 1 単語が試行ごとにランダムに選択され, 文字刺激として 3 回連続して呈示を行った。1 回目および 2 回目の呈示時には, 文字に対応する音声も同時に呈示されるが, 3 回目の呈示時には音声は呈示されない。Fig. 1 に 1 試行ごとの呈示の流れを示す。

被験者には, 3 回目の文字呈示にタイミングを合わせて, 対応する音声を想起するように求めた (想起条件)。また, 比較のために, 3 回目の文字呈示時に想起を行わせない条件でも計測を行った (コントロール条件)。また, 両条件に共通して, 呈示区間において文字刺激と対応しない音声刺激が呈示された場合には, ボタンで応答するように教示した。

文字刺激および音声刺激の持続時間は約 800 ms, 刺激間隔 (Inter-stimulus interval : ISI) は 500 ms とした。試行開始の合図音と 1 回目の刺激呈示までの間隔は 500 ms, 試行間隔は眼球運動のアーティファクト除去方法に応じて,

* Estimation of features extracted from brain magnetic field data about discrimination of imagined sounds. by UZAWA, Shihomi (Kobe Univ/AIST), TAKIGUCHI, Tetsuya, ARIKI, Yasuo (Kobe Univ.), NAKAGAWA, Seiji (Chiba Univ/AIST).

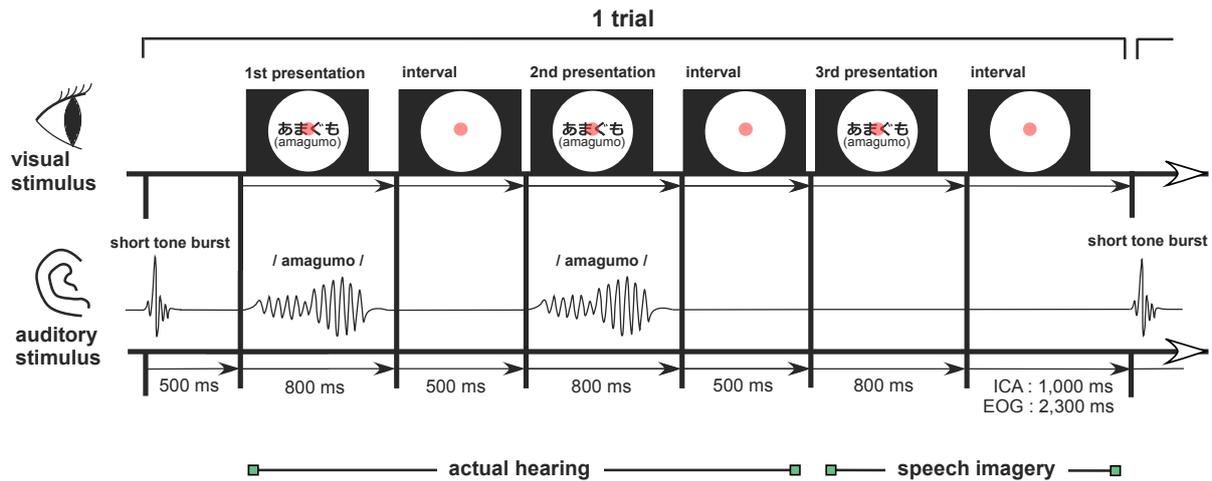


Fig. 1 Schematic diagram of task.

1000 ms もしくは 2300 ms とした。また、用いた音声刺激は、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を利用した。

2.2 脳磁界計測

聴覚健常者 8 名 (男性 7 名, 女性 1 名, 20-40 歳) を被験者とした。

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 変換を行った。得られた生データに対して、独立成分分析 (independent component analysis : ICA) を適用し、眼球運動に伴うアーティファクトを除去した。

3 解析方法

3.1 CWT による特徴量抽出

3 回目の呈示区間における生データ波形に対して音声想起に伴う反応成分を抽出した。文字刺激呈示から 0-1200 ms の区間の左右側頭部に相当する 44 チャンネルの脳磁界波形に対し、以下の式に従って連続ウェーブレット変換 (Continuous Wavelet Transform: CWT) による特徴量抽出を行った。

$$CWT(b, a) = \frac{1}{\sqrt{a}} \int f(t) \psi\left(\frac{t-b}{a}\right) dt \quad (1)$$

$f(t)$ は時系列脳磁界波形とする。 $\psi(t)$ はウェーブレット関数 (マザーウェーブレット) であり、本稿では複素モルレーウェーブレットを用いた。

a, b はマザーウェーブレットに係るパラメータであり、それぞれスケールと時間シフトを表している。

各被験者データにおいて変換を行い、得られる特徴量は 2.5ms ごとに 0-1200 ms の 480 次元、1 Hz ごとに 1-50 Hz の 50 次元、44 チャンネルによる 44 次元から構成される。特徴量抽出の過程を Fig. 2 に示す。抽出された特徴量は、主成分分析 (Principal Component Analysis: PCA) による周波数次元の削減を行い、累積寄与率が 80% を満たす第一主成分のみを採用した。

3.2 SVM による識別

解析には、50 試行分の脳磁界データを使用した。そのうち、ある 1 試行を評価データ、残りの 49 試行を学習データとするデータセットを、評価データを入れ替えることで 50 組作成し、ガウシアンカーネルによる非線形マルチクラス SVM を用いて、50-fold Cross Validation を行った。学習データセットには、1 試行ごとに特徴量抽出によって得られた 21120 次元 (時間: 480 次元, チャンネル: 44 次元, 周波数: 1 次元) の特徴量ベクトルと想起音声ラベルの組による 3 つのサンプルを用いた。また、反応の時間のずれを考慮し、0-10 ms だけ時間を遅らせた特徴量ベクトルも同様に学習に用いた。

生成した学習モデルに対応する評価データセットによって各単語データにおける正答率を得たのち、解析に用いたモデルすべての平均正答率を想起音声ごとと 3 単語全体それぞれに対して算出した。

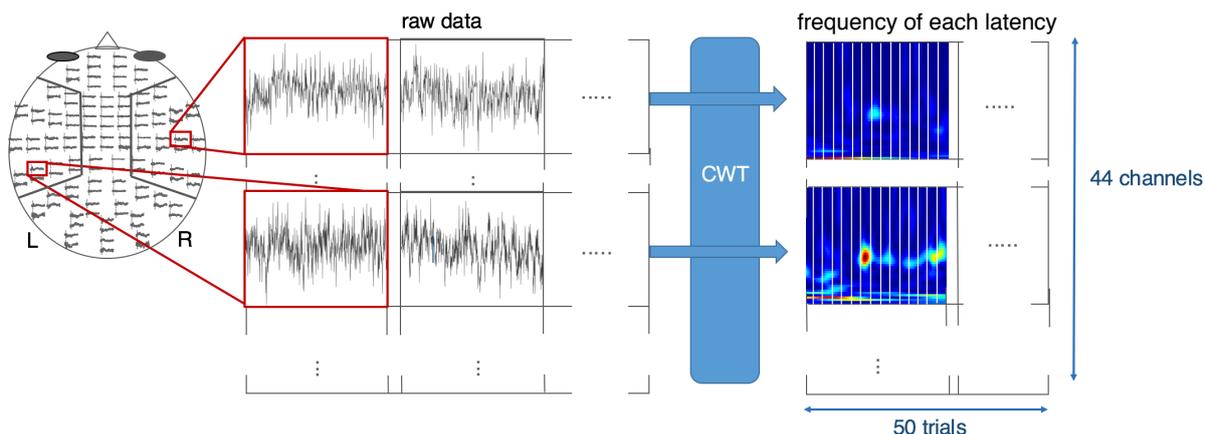


Fig. 2 Process of feature extraction.

さらに、3単語がいずれも正解に分類されたときの学習モデルについて、分離超平面の構成に使用したサポートベクターから特徴量次元ごとの重みを算出した。本稿では、one versus one による多値分類を行ったため、3組のクラスのペアに応じた重みを得た。

4 結果

Table 1 に想起音声ごとの脳磁界波形の識別における正答率を示す。全被験者で平均して74.0%の正答率が得られた。想起音声ごとの識別結果から、「うらない」が他の言語に比べて高い精度が得られた。

また、想起音声のペア3組の特徴量次元重みの平均値を Table 2 に示す。さらに、それら重みを潜時について平均した値を、奇数チャンネルと偶数チャンネルにわけて Fig. 3 に示す。また、チャンネルについて平均した値を Fig. 4 に示す。潜時については、(緯線方向の空間差分)の重みは潜時に応じておおよそ一定であるのに対し、奇数チャンネル(経線方向の空間差分)の重みは610 msで最も大きくなった。また、チャンネルについてはやや右側側頭部の重みが大きいことがわかった。

5 考察

加算平均を施していない raw data 波形に対して交差検定を行った結果、全体として74.0%の精度が得られた。音声聴取時の誘発脳磁界波形は、聴取する単語によって変化することが知られている [7]。Table 1 は、自発脳磁界や環境雑音のために S/N 比が低いデータについても、学習モデル上でチャンネルや潜時に単語の違いが反映されることを示している。一方、想起音声ごとに正

Table 1 Discrimination accuracy of the imagined sound.

| Sub. | Word | | | Ave. |
|------|---------|--------|--------|------|
| | amagumo | ibento | uranai | |
| 1 | 62.0 | 58.0 | 98.0 | 72.7 |
| 2 | 86.0 | 76.0 | 94.0 | 85.3 |
| 3 | 58.0 | 48.0 | 98.0 | 68.0 |
| 4 | 64.0 | 64.0 | 100.0 | 76.0 |
| 5 | 46.0 | 36.0 | 98.0 | 60.0 |
| 6 | 64.0 | 64.0 | 100.0 | 76.0 |
| 7 | 56.0 | 52.0 | 100.0 | 69.3 |
| 8 | 74.0 | 80.0 | 100.0 | 84.7 |
| Ave. | 63.8 | 59.8 | 98.5 | 74.0 |

答率を見たとき、「うらない」が他の2単語に比べて大きく、「いべんと」が最も小さい。Table 2 から学習モデルを生成する際の(いべんと, うらない)の重みに比べて、(あまぐも, いべんと)と(あまぐも, うらない)の重みが大きいことがわかるが、このことは(いべんと, うらない)の分離がうまく機能せず、結果としてうらないの正答率が上昇したことを示唆している。また、本稿では、CWTとPCAを用いて特徴量抽出を行ったが、CWTは時間ごとに周波数を捉える反面、情報が膨大になる可能性があるため、過学習に陥った可能性は否定できない。次元削減処理についての検討を慎重に行う必要があると思われる。

Fig. 1 より、右半球のチャンネルの重みがやや大きいことがわかる。この結果は、先行研究 [3] と同様に、右側がやや優位であることを示唆している。また、奇数チャンネルでは潜時に伴って重みが

Table 2 Average of weight of feature vector for each imagined sound in 50 trials.

| Sub. | Pair | | |
|------|------------|------------|------------|
| | (ama, ibe) | (ibe, ura) | (ama, ura) |
| 1 | 36.7 | 1.13 | 37.7 |
| 2 | 34.3 | 0.98 | 35.1 |
| 3 | 34.0 | 4.18 | 37.4 |
| 4 | 34.2 | 0.04 | 34.3 |
| 5 | 37.8 | 3.68 | 37.0 |
| 6 | 31.3 | 4.60 | 32.5 |
| 7 | 30.9 | 3.61 | 31.4 |
| 8 | 21.3 | 9.26 | 26.0 |
| Ave. | 32.58 | 3.43 | 33.9 |

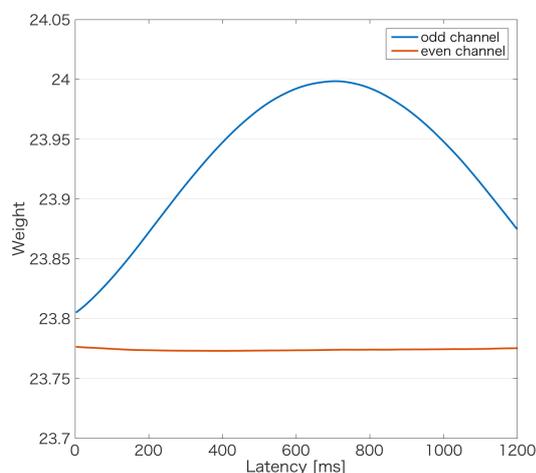


Fig. 3 Average of weight of the latency in 50 trials.

変化し、610ms で最大の値を示した。奇数チャンネルは経線方向の磁場の差分、すなわち緯線方向（前後方向）の脳内電流を捉えている。磁場パターンを考慮すると、後ろ向きの神経電流成分の潜時610ms 付近に、音声の違いを示す特徴量が含まれていると考えられる。

6 おわりに

脳磁界計測の生波形に対して、CWT と SVM を用いて、交差検定による想起音声の違いを分類した。本実験での想起音声の違いを判別するにあたって有効な特徴量は、右側側頭部チャンネルと潜時約 600ms 付近がやや有効であることが示唆された。

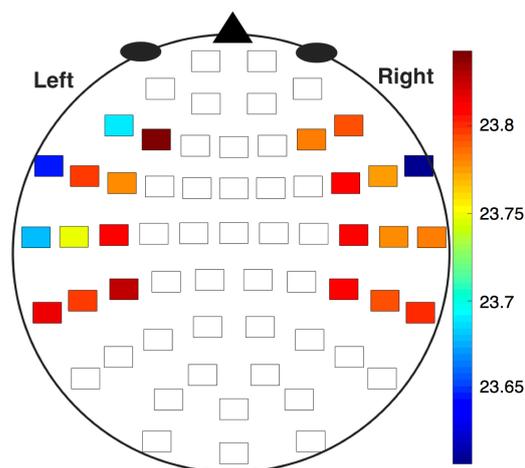


Fig. 4 Average of weight of the channel in 50 trials.

謝辞 本研究の一部は科学研究費補助金（26282130, 26560320, 25282053）の援助を受けて実施された。

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] L. Jäncke *et al.*, “‘Hearing’ syllables by ‘seeing’ visual stimuli,” *European Journal of Neuroscience*, Vol. 19, pp. 2603-2608, 2004.
- [3] M. Hoshiyama *et al.*, “Hearing the sound of silence: a magnetoencephalographic study,” *NeuroReport*, Vol. 12, pp. 1097-1102, 2001.
- [4] 宇澤ら, “音声想起による誘発脳磁界の計測,” 日本音響学会 2016 年春季研究発表会, 2016.
- [5] Ayad G. Baziyad *et al.*, “A Study and Performance Analysis of Three Paradigms of Wavelet Coefficients Combinations in Three-class motor imagery based BCI,” in *Proc. ISMS*, pp. 201-205, 2014.
- [6] G. Pfurtscheller *et al.*, “Event-related EEG/MEG synchronization and desynchronization: basic principles,” *Clinical neurophysiology*, vol. 110, pp. 1842-1857, 1999.
- [7] M. Bourguignon *et al.*, “The pace of prosodic phrasing couples the listener’s cortex to the reader’s voice,” *Human Brain Mapping*, Vol. 34, pp. 314-326, 2013.