脳磁界データによる音声の識別 -想起時と聴取時の比較- *

☆矢野彩緒里, 高島遼一, 滝口哲也, 有木康雄 (神戸大),
添田喜治 (産総研), 中川誠司 (千葉大/産総研)

1 はじめに

身体障害者の生活支援の手段として，音声認識や行動認識を用いた機械制御・意思伝達技術が用いられている．しかし，発話や身振りでのコミュニケーションが困難な障害者は，音声認識技術を有効に活用することができない．そこで，脳活動を用いて機械制御をおこなうブレイン・コンピュータ・インターフェイス (BCI) の利用が期待される．

従来の BCI 開発では，注意を向けた低頻度刺激に対してのみ出現する誘発反応 (P300) によって，限定的な意思伝達をおこなう “P300 スペラー型 [1]” の開発例が多い．しかし，P300 スペラー型では予め用意された選択肢の中から使用者の意思が判別される仕組みとなっており，自由度の高い意思伝達はできない．そこで我々は，ユーザの意思をより汎用的に認識する BCI の実現に向け，脳磁界データから想起音声の識別を試みた．

我々はこれまで，脳磁界計測データを用いて，音声聴取時との比較から音声想起時の時空間特性の検証をおこなってきた．想起音声のエンベロープと想起により誘発された加算平均脳磁界反応の相関を調べたところ，想起時には音声聴取時と同レベルの相関は得られなかったものの，時間波形上で聴覚野周辺の緩やかな活動 が認められた [2]．また，想起音声の単語識別において，脳磁界データのチャネル・周波数・時間における多次元情報を損なわないよう，テンソル分解により低次元な特徴量を抽出し，学習モデルの評価を行った [3]． さらに，空間的な特徴を損なわない識別についても検討を行ったが，実用上十分な識別精度が得られとは言い難い [4]．音声想起時の脳活動には音声聴取時に比べて明らかな音響特徴の鈍化・欠落が生じており，単語を識別するのは容易ではないと思われる．

本研究では，脳磁界データにおける想起音声識別の精度向上を図る上で有用な情報を得るため，同一被験者の音声聴取時と音声想起時の脳活動の識別を試みた．識別における聴取時と想起時の関連性を検討するため，双方の識別タスクにおいて Wavelet 特徴量，Common Spatial Pattern で特徴量抽出をし，サポートベクターマシーンにより識別を行い，各手法における識別精度を比較した．

2 脳磁界計測と実験内容

3 名の右利きの聴覚健常者に対し，3 パターンの単語音声 (“あまぐも”，“いべんと”，“うらない”) を用いて実験をおこなった．3 単語のうち 1 単語がランダムに選択され，文字刺激として 3 回呈示をおこなった．1 回目と 2 回目には同時に音声刺激も呈示した．被験者には，1 回目，2 回目の単語音声を聴取した後，3 回目の文字呈示のタイミングで聴取した音声を想起するよう求めた．1 試行の概要を Fig. 1 に示す [2].

音声刺激には，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 (fto) を利用した．刺激呈示時間は 800 ms であり，解析対象の脳磁界データの解析時間は，聴取時は聴取の 200ms 前，想起時は想起開始の合図の 200ms 前から 1200 ms 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 - 122^{TM}: 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 変換をおこなった．同時に被験者 1〜 3 については眼電図 (Electrooculogram : EOG) の測定もおこない， EOG が 5 mV を超えた際の脳磁界データは学習・テストデータから除外した．

3 特徴量と識別手法

本実験では単一試行波形に対して独立成分分析 (Independent Component Analysis : ICA) により眼電除去を行った．左右側頭部聴覚野を覆う計 36 チャネルに対して，文字刺激提示 200ms 前か

* Sound classification of brain magnetic fields -Comparison of imagery and hearing-, YANO, Saori, TAKASHIMA, Ryoichi, TAKIGUCHI, Tetsuya, ARIKI, Yasuo (Kobe Univ.), SOETA, Yoshiharu (AIST), NAKAGAWA, Seiji (Chiba Univ./AIST).

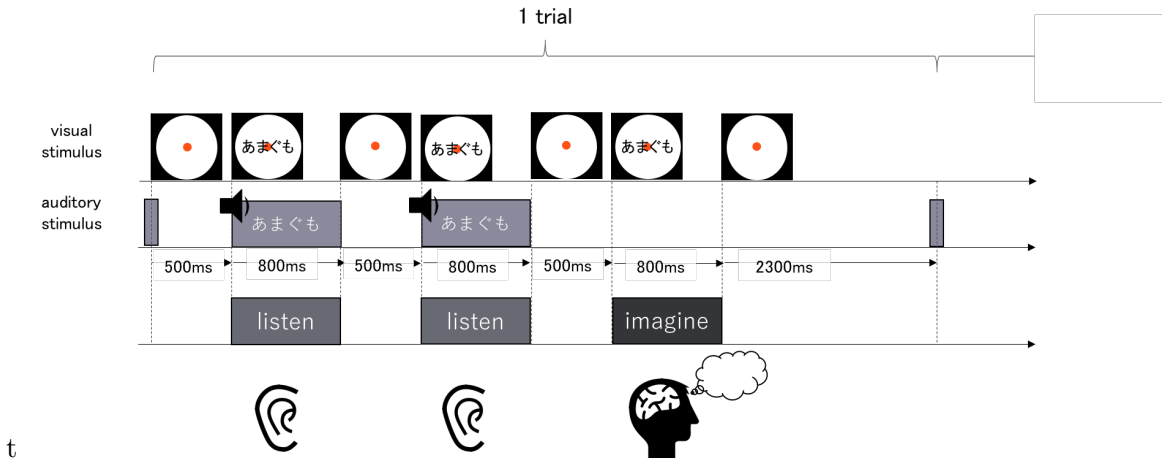

Fig. 1　Schematic diagram of the task.

ら提示後 1000 ms の区間，計 1200ms 間の単一試行波形または加算波形に対して，以下の特徴量抽出をそれぞれ行った．

3.1　Common Spatial Pattern

Common Spatial Pattern (CSP) は運動想起などで高精度を示す特徴量抽出法であり，空間重みを用いた変換を行う [5]．CSP は，以下の式 (1) により得られる．

$$\mathbf{x}_{csp} = \mathbf{W}^{\mathbf{T}}\mathbf{x} \tag{1}$$

ここで，$\mathbf{x}$ はもとの信号，$\mathbf{W} \in \mathbb{R}^{N\times L}$ は空間フィルタ，N はチャネル数，L は CSP のコンポーネント数を示す．

マルチクラスの CSP は式 (2) を満たす $\mathbf{W} \in \mathbb{R}^{N\times N}$ を求めることにより得られる [6]．

$$\mathbf{W}^T\mathbf{R}_{\mathbf{x}|C_i}\mathbf{W} = \mathbf{D}_C, \quad i = 1, ..., M \tag{2}$$

この式における $\mathbf{R}_{\mathbf{x}|C_i}$ はクラス i における分散共分散行列，M はクラス数である． また，$\boldsymbol{\lambda}_i = diag\{\mathbf{W}^T\mathbf{R}_{\mathbf{x}|C_i}\mathbf{W}\}$ とし，$\lambda_{i,j} = max\{\lambda_{i,j}, 1/(1+(M+1)^2)\lambda_{i,j}/(1-\lambda_{i,j}))$, $j = 1, ..., N$ とした際に各クラスにおいて固有値 $\lambda_{i,j}$ の降順 L/M 個に対応する固有ベクトル $\mathbf{W}_j$ を式 (1) の変換行列とする．本実験では，1-50Hz に対しバンド幅を 4Hz，周波数シフトを 2Hz として周波数フィルタをかけ，それぞれに対して CSP を行い，入力特徴量とした．

3.2　連続ウェーブレット変換

連続ウェーブレット変換 (Continuous Wavelet Transform : CWT) による時間周波数特徴量の抽出を行った．CWT 関数 W は以下の式 (3) で示される．

$$W(a,b) = \frac{1}{\sqrt{a}}\int x(t)\psi(\frac{t-b}{a})dt \tag{3}$$

ここで，$x(t)$ は脳磁界の時系列波形である．$\psi(t)$ はマザーウェーブレットであり，本実験では Morlet ウェーブレットを用いた．a はスケール，b は時間シフトを表すマザーウェーブレットのパラメータである．

CWT 関数 W には，1 チャネルごとに周波数方向に 1-50 Hz の 50 次元，時間フレームとして 480 次元が得られる．本実験では特徴量削減のため，フレーム長 100ms，フレームシフト 50ms とし，時間フレームを 23 次元に圧縮した．

3.3　SVM による識別

音声聴取時または想起時の単一試行波形より得られた CSP，CWP を特徴量とし，サポートベクターマシーン (Support Vector Machine : SVM) により識別を行う．データ数は被験者ごとに異なり，聴取時については平均 635 分散 1115，想起時については平均 317.5 分散 278.8 である．全データに対し 8 割を学習データ，2 割をテストデータとして 5 fold cross validation を行い，識別精度の平均を算出した．また，学習時にはグリッドサーチから，分類精度の高いものをハイパーパラメータとして決定した．

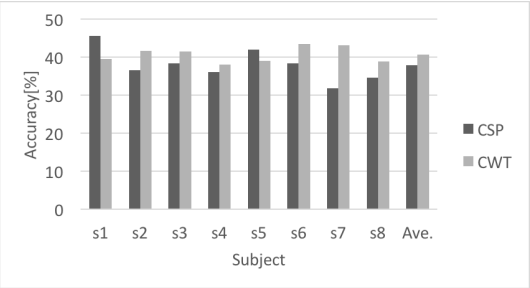

Fig. 2　Accuracy of actual hearing.

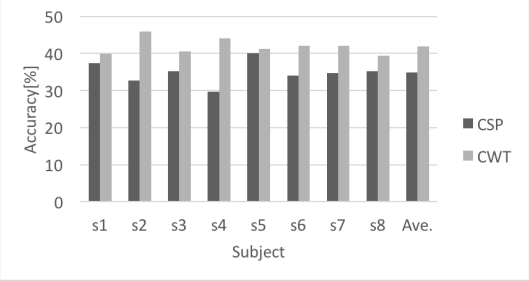

Fig. 3　Accuracy of speech imagery.

4　実験結果と考察

4.1　特徴量の比較

まずはじめに，音声聴取時の被験者ごとの識別精度を Fig. 2 に示す．CSP 特徴量としては 36 チャネルに対して変換を行いコンポーネント数 10 としたもの，CWT 特徴量としては 1 チャネルごとに分類を行い識別精度の最も高かったチャネルについての結果を示している．

被験者 1 の CSP が 45.6%と最も高い識別率を示した．しかし，被験者 1, 5 を除いては CSP よりも CWT の識別率が高く，全被験者の平均としては，CSP が 37.9%，CWT が 40.6%を示した．また，CWT においては全被験者において識別率 38%を超えた．

次に，音声想起時の被験者ごとの識別精度を Fig. 3 に示す．特徴量の条件は聴取時と同様である．被験者 2 の CWT が 45.9%の識別率を示し，被験者 1 と被験者 8 を除き識別率 40% を超えた．全被験者の平均としては CSP が 34.9%，CWT が 41.9 %と，聴取時よりも CWT が強い優位性を示した．

また，聴取時と想起時を比較すると，被験者 1, 5 など聴取時の CSP の識別率が高い被験者は想起時においても他の被験者より高い識別度を示した．しかし，想起時においてはいずれも CWT の識別精度を下回った．

4.2　チャネルごとの識別率の比較

音声聴取時よりも想起時に CWT の CSP に対する優位性が顕著になる理由として，チャネル選択が不適切であるということが予測される．そこで，次に CWT 特徴量を用いて音声聴取時，想起時のチャネルごとの識別率について比較した．本実験は単一試行波形と，同期加算をすることによりノイズを軽減した 5 回加算波形の 2 波形に対して行った．被験者 1-3 についてチャネルごとの識別率の分布を表したものを Fig. 4 に示す．N は脳磁界波形の加算回数を示している．

被験者 1-3 において，単一試行波形においても 5 回加算波形においても，聴取音声の識別率は主に右半球において高い値を示した．これは右側頭部の聴覚野で音声の識別が行われていることを意味する．また，想起音声については被験者間で識別率の高いチャネルが大きく異なっており，音声想起に起因する脳活動が被験者間で大きく異なることが予測される．

Fig. 4 において，chance rate (33.3%) よりも識別率が低いチャネルを濃いグレーで示している．被験者 1 の聴取音声識別において，5 回加算波形の識別率は全チャネルにおいて chance rate を上回った．しかし，想起音声については chance rate を下回るチャネルが多く存在しており，側頭部 36 チャネル全てを識別に使った際の精度低下の原因と考えられる．

また被験者 1 については聴取音声・想起音声において識別率が高いチャネルが異なっているが，被験者 2, 3 において聴取音声と想起音声の識別率に強い相関が見られる．このことから，被験者 2, 3 においては音声想起を行う際，音声聴取と類似した脳活動が行われているのではないかと推測できる．

以上の結果から，想起音声の識別率向上に向けて次の方針を立てる．まず，適切なチャネル選択が必要である．本実験にて 36 チャネルの信号を用いた CSP の識別精度が 1 チャネルの CWT を下回ってしまったが，識別率が高くなるようなチャネルを上手く選択することにより多チャネル計測の利点を生かした識別が行えるのではないかと考える．また，被験者 2, 3 のように聴取時と想起時の識別に相関が見られるものについては，想起音声の識別をする際の学習に何らか

の形で聴取時の脳磁界データを用いることで識別精度が向上する可能性があると考える．今後，以上の条件を満たすような学習モデルを検討する必要がある．

5　おわりに

本稿では，音声聴取時と想起時の脳磁界データに対し，連続ウェーブレット変換，Common Spatial Pattern を特徴量とした識別を行った．

想起音声の識別に対し，連続ウェーブレット変換が Common Spatial Pattern の識別率を大きく上回った．またその結果を受け，チャネルごとの識別率について音声聴取時と想起時の比較を行った．

識別精度向上に向けた今後の課題としては，チャネル選択を適切に行い，想起音声識別の学習データに音声聴取時の脳磁界データを用いるような学習モデルを検討していきたい．

謝辞　本研究の一部は，JSPS 科研費 JP18K19820 の助成を受けたものである．

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] S. Uzawa *et al.*, “Spatiotemporal Properties of Magnetic Fields Induced by Auditory Speech Sound Imagery and Perception,” IEEE EMBC2017, pp. 2542-2545.

[3] 宇澤ら, “脳磁界データによる想起音声の識別 -次元数削減による精度向上の検討-,” 日本音響学会 2017 年 秋季研究発表会, pp. 337-340.

[4] 矢野ら, “脳磁界データの空間的特徴を考慮した想起音声の識別,” 日本音響学会 2018 年 秋季研究発表会, pp. 337-340.

[5] H. Ramoser *et al.*, “Optimal spatial filtering of single trial EEG during imagined hand movement,” IEEE Transactions on Rehabilitation Engineering, vol. 8, no. 4, pp. 441-446, 2000.

[6] M. Grosse-Wentrup *et al.*, “Multiclass common spatial patterns and information theoretic feature extraction,” IEEE Transactions on Biomedical Engineering, Vol 55, no. 8, pp. 1991-1999, 2008.

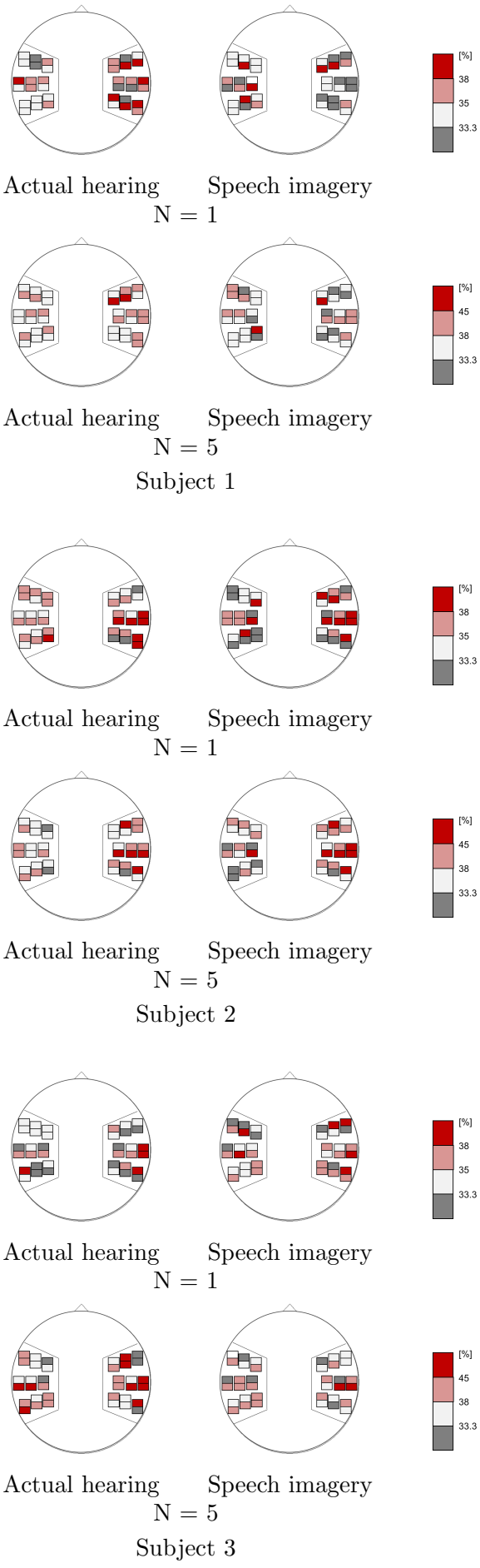

Fig. 4　Accuracy mapping of each channel.