

[ポスター講演] 日本語音声想起に伴う脳磁界データの判別と特徴量推定

宇澤志保美^{†,††} 滝口 哲也[†] 有木 康雄[†] 中川 誠司^{†††,††}

[†] 神戸大学大学院システム情報学研究科 〒657-8501 兵庫県神戸市灘区六甲台町 1-1

^{††} 国立研究開発法人産業技術総合研究所 〒563-8577 大阪府池田市緑丘 1-8-31

^{†††} 千葉大学フロンティア医工学センター 〒263-8522 千葉県千葉市稲毛区弥生町 1-33

E-mail: †uzawa@me.cs.scitec.kobe-u.ac.jp, ††{takigu,ariki}@kobe-u.ac.jp, †††s.nakagawa@chiba-u.ac.jp

あらまし 本研究では、日本語音声の想起活動に伴う脳磁界計測データから想起音声の判別と、判別に関連する識別特徴量の推定を行う。近年、脳活動を機械制御信号や意思伝達手段として活用するブレインコンピュータインターフェース (BCI) の開発研究が進められている。しかし、意思伝達型 BCI システムの多くは、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) を用いており、あらかじめ用意された選択肢の中からの判別することしかできない。より汎用性のある BCI の開発のためには、ヒトの脳活動から意思そのものを解読できることが望ましい。本研究ではまず、連続ウェーブレット変換 (Continuous Wavelet Transform: CWT) を用いて日本語音声想起時の脳磁界の時間周波数特徴量を計算した。次に、時間周波数特徴量に対して非負値テンソル分解 (Non-negative Tensor Factorization: NTF) を適用し、多次元構造を考慮した低次元の特徴量を抽出した。最後に、得られた低次元特徴量ベクトルを用いて、想起音声の識別を行った。

キーワード 脳磁界計測, 音声想起, 非負値テンソル分解

Discrimination and Feature Estimation of Brain Magnetic Field Data Associated with Japanese Speech Sound Imagery

Shihomi UZAWA^{†,††}, Tetsuya TAKIGUCHI[†], Yasuo ARIKI[†], and Seiji NAKAGAWA^{†††,††}

[†] Graduate School of System Informatics, Kobe University Rokkodai 1-1, Nada-ku, Kobe-shi, Hyogo,
657-8501 Japan

^{††} National Institute of Advanced Industrial Science and Technology (AIST) Midorigaoka 1-8-31,
Ikeda-shi, Osaka, 563-8577 Japan

^{†††} Center for Frontier Medical Engineering, Chiba University Yayoi-cho 1-33, Inage-ku, Chiba-shi, Chiba
263-8522 Japan

E-mail: †uzawa@me.cs.scitec.kobe-u.ac.jp, ††{takigu,ariki}@kobe-u.ac.jp, †††s.nakagawa@chiba-u.ac.jp

Abstract Brain computer interface (BCI) technologies, which enable direct communication between the brain and external devices, have been developed. However, most of the BCI systems that have been developed are the “P300-speller” type, which can only detect objects that users direct his/her attention at. With such BCI systems, users can only choose something from among what was prepared in advance. To develop more versatile BCI systems that can detect a user’s intention or thoughts, the brain responses associated with verbal imagery need to be clarified. In this study, first, the time-frequency features were extracted from the brain magnetic field data associated with Japanese speech sound imagery using continuous Wavelet Transform (CWT). Second, the low-dimensional features were extracted from the time-frequency features using non-negative tensor factorization (NTD). Finally, the imagined sounds were classified by using the extracted low-dimensional feature vectors.

Key words magnetoencephalography (MEG), speech sound imagery, non-negative tensor decomposition

1. はじめに

近年、脳活動を意思伝達手段として活用する試みとして、ブレイン・コンピュータ・インターフェース (BCI) の開発研究が盛んに行われている。中でも、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、間接的に意思伝達を行う報告は多い。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、ヒトの脳活動から意思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、活動の空間特性はバラツキが大きく、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、脳磁界計測データを用いて、観測のしやすい音声聴取時の脳活動と比較することにより、音声想起時の時間空間特性の検証を行ってきた。音声想起活動が音声聴取時のように音声を脳内で生成・聴取していると仮定し、想起音声の振幅エンベロープと想起によって誘発された加算平均脳磁界反応の相関を推定した。その結果、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ことから [4]、次は振幅エンベロープ以外のパラメータとして時間周波数特徴量に注目した。認知活動に伴う特定の周波数帯域の変動を報告した例は多く [5]、音声想起でも同様な変動が観察される可能性は高い。また、微細な特徴量の変化を捉えるために機械学習を用いた解析を行った [6] が、想起音声や被験者に依存しない頑健な特徴量を抽出できたとは言い難い。

音声想起の脳磁界データの時間周波数特徴は、周波数や時間、さらにはチャンネルなどの多次元情報を持ち、特徴量ベクトルの次元数が非常に大きくなりがちである。特徴量次元の多さは、偏った識別結果を生み、過学習や計算コストの増大の要因となる。また、微弱な音声想起の脳反応は、ノイズの影響を大きく受けてしまう。そのため、適切な次元削減処理は、正確な識別結果をもたらすために不可欠である。本研究では、識別精度の偏った結果の回避、および識別精度の向上を目的として、適切な次元削減処理を施し、脳磁界データから想起音声の識別を試みた。連続ウェーブレット変換と非負値テンソル分解によって次元削減を施した特徴量の抽出を行い、サポートベクタマシンにより想起音声の識別を行った。

2. 計測方法

2.1 刺激呈示およびタスク

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。このとき、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。図 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、1000 ms もしくは 2300 ms とした。また、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を利用した。

2.2 脳磁界計測

聴覚健常者 4 名 (男性 3 名, 女性 1 名, 20-40 歳) のデータを解析対象とした。

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

CWT と NTF を用いて特徴量抽出を行い、識別にはサポートベクタマシン (Support vector machine: SVM) を用いた。図 2 に解析の過程を示す。

3.1 CWT による特徴量抽出

3 回目の呈示区間の脳磁界データから、音声想起に伴う反応成分を抽出した。文字刺激呈示から 0-800 ms の区間の左右側頭部に相当する 44 チャンルの脳磁界波形に対して、以下の式に従って CWT による時間周波数特徴量の抽出を行った。

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

$f(t)$ 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。

脳磁界波形には、加算平均の施されていない Raw 波形、Raw データからランダムに 5 もしくは 10 個のエポックを取り出した 2 種類の誘発波形の計 3 種類を用いた。誘発波形は、CWT によって Raw 波形の時間周波数特徴を得たのち、学習データについては $80C_{5or10}$ 通りの組み合わせのうちの 80 個、評価データについては $20C_{5or10}$ 通りの組み合わせのうちの 20 個の時間周波数特徴求めて、それぞれデータセットとして用いた。

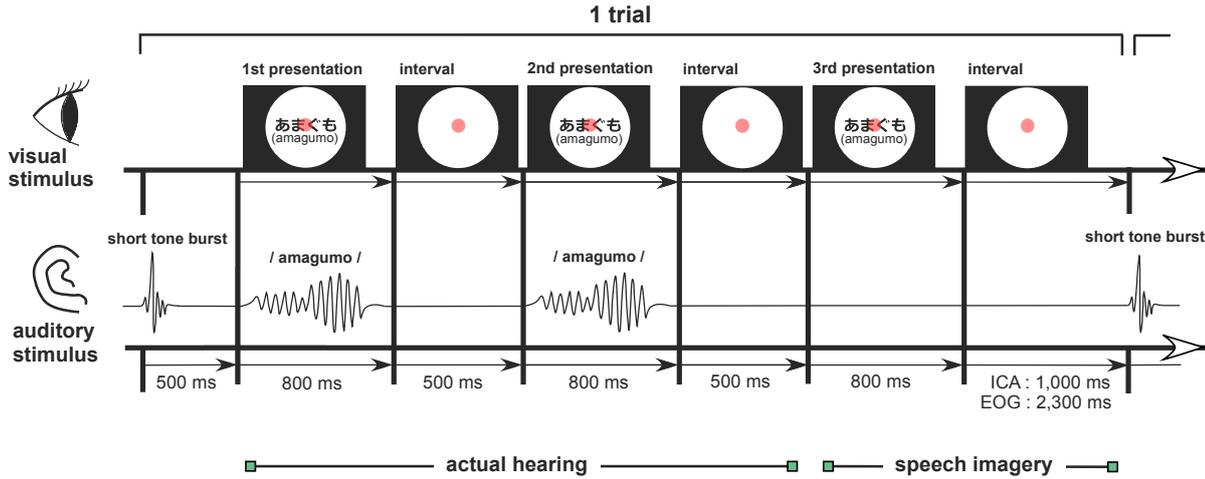


図1 刺激呈示タスクの流れ
Fig.1 Schematic diagram of the task.

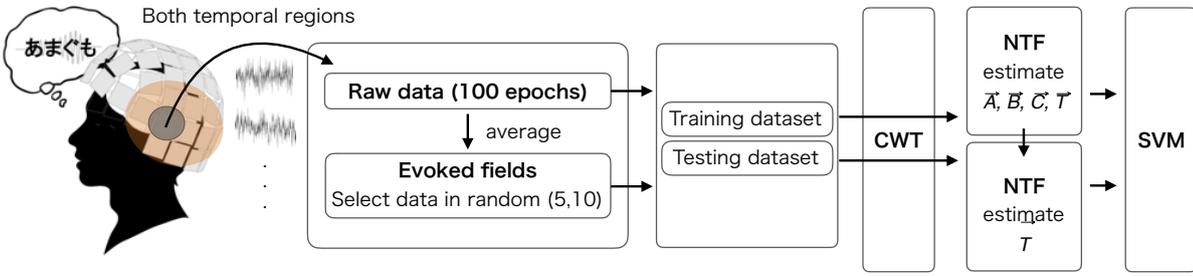


図2 特徴量抽出および識別の流れ
Fig.2 Process of the feature extraction.

得られる特徴量は、2.5ms ごとに 0-800 ms の 320 次元、1 Hz ごとに 1-50 Hz の 50 次元、44 チャンネルによる 44 次元から構成されている。

3.2 NFT による特徴量抽出

時間周波数特徴に変換した脳磁界計測データは、周波数、時間、チャンネルに加えて、計測試行の数を含めた 4 次元配列によって表現できる。これらの特徴量を識別学習や評価に用いるためには、各次元を考慮しながらベクトル化する必要があるが、多次元情報をベクトル化すると特徴量ベクトルの次元数が非常に大きくなり、次元の呪いや計算コスト増大の問題、そして不必要な情報を含有することによる識別率の低下が懸念される。また、ベクトル化により、(周波数) × (時間) × (チャンネル) × (試行) といったデータ構造の情報が失わないような特徴量ベクトルを作成することが望ましい。そのため、脳磁界計測データの CWT による時間周波数特徴量に対し、NTF を用いて多次元構造を考慮した低次元の特徴量を抽出した。

周波数、時間、チャンネル、試行数の次元を持つ 4 階のテンソル $\underline{X} = \{x_{i,j,k,l}\} \in \mathbb{R}_+^{I \times J \times K \times L}$ に対する NFT は以下の式で表される。[7], [8]

$$\underline{X} \approx \sum_{r=1}^R \mathbf{a}_r \circ \mathbf{b}_r \circ \mathbf{c}_r \circ \mathbf{t}_r \quad (2)$$

$\mathbf{A} = \{a_{i,r}\} = [\mathbf{a}_1 \dots \mathbf{a}_R] \in \mathbb{R}_+^{I \times R}$, $\mathbf{B} = \{b_{j,r}\} = [\mathbf{b}_1 \dots \mathbf{b}_R] \in \mathbb{R}_+^{J \times R}$, $\mathbf{C} = \{c_{k,r}\} = [\mathbf{c}_1 \dots \mathbf{c}_R] \in \mathbb{R}_+^{K \times R}$, $\mathbf{T} = \{t_{l,r}\} = [\mathbf{t}_1 \dots \mathbf{t}_R] \in \mathbb{R}_+^{L \times R}$ は、それぞれ周波数、時間、チャンネル、試行に関する基底行列である。また、テンソル \underline{X} の各要素は以下の式で表される。

$$x_{i,j,k,l} \approx \sum_{r=1}^R a_{i,r} b_{j,r} c_{k,r} t_{l,r} \quad (3)$$

“ \circ ” はベクトルの直積を表す。基底行列 \mathbf{A} , \mathbf{B} , \mathbf{C} , \mathbf{T} は、式 (2) の左辺と右辺の差の Frobenius ノルムを最小化するような最適化問題を解くことによって求めた。基底行列 \mathbf{A} , \mathbf{B} , \mathbf{C} , \mathbf{T} を順番に、非負値行列因子分解の乗法更新式を基にした更新式 [7] を用いて、収束するまで繰り返し更新を行うことにより、局所最適解を得た。なお、基底行列のスケールの任意性から、 \mathbf{A} , \mathbf{B} , \mathbf{C} の各列ベクトルの L2 ノルムが 1 になるように正規化した。

式 (2) に示した NFT は、テンソル \underline{X} を R 個の周波数、時間、チャンネルの次元を持つ 3 階のランク 1 テンソル $\mathbf{a}_r \circ \mathbf{b}_r \circ \mathbf{c}_r$

表 1 想起音声データの識別精度

Table 1 Discrimination accuracy of the imagined sounds.

Sub.	Raw (N=1)				Average (N=5)				Average (N=10)			
	w1	w2	w3	Ave.	w1	w2	w3	Ave.	w1	w2	w3	Ave.
1	53.3	36.7	16.7	35.6	38.0	25.4	47.3	36.9	44.1	28.3	33.6	35.3
2	31.7	35.0	28.3	31.7	48.4	33.3	27.5	36.4	40.0	38.8	30.0	36.3
3	18.3	48.3	40.0	35.5	50.5	43.5	10.4	34.8	61.4	22.6	21.5	35.2
4	63.3	30.0	16.7	36.7	39.4	33.5	40.1	37.7	28.9	48.6	34.3	37.3
Ave.	41.7	37.5	25.4	34.9	44.1	33.9	31.3	36.4	43.6	34.6	29.9	36.0

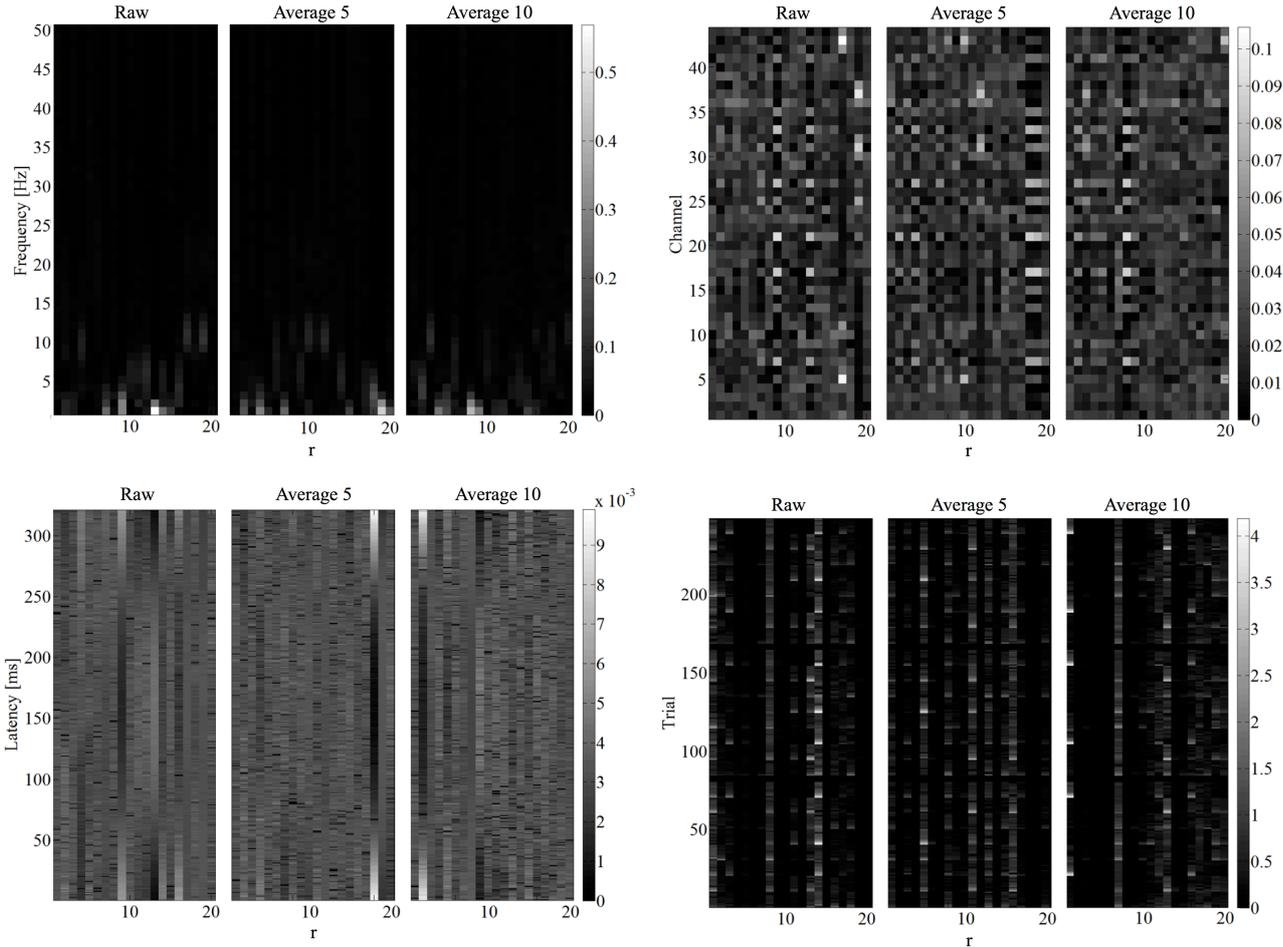


図 3 被験者 4 において、学習データを用いて推定した NTF の基底行列。 **A** (周波数: 左上), **B** (潜時: 左下), **C** (チャンネル: 右上), **T** (試行: 右下).

Fig. 3 The basis matrix of NTF by using training data in subject 4. **A** (frequency: top left), **B** (latency: bottom left), **C** (channel: top right), **T** (trial: bottom right).

に、それぞれ係数 $t_{i,r}$ を乗じて足し合わせたモデルと解釈できる。各ランク 1 テンソルが個々の脳活動を表現すると仮定すると、基底行列 **T** の行ベクトルがそれらの強度を表すと考えられる。

3.3 SVM による識別

特徴量の抽出及び識別には、各被験者において、単語ごとに 100 個の脳磁界波形を用いた。そのうち、20 試行を評価データ、残りの 80 試行を学習データとした。学習及び評価データについては、Raw 波形に対しては評価試行を入れ替え、誘発波

形に対しては加算の組み合わせを変えることにより、各波形において 8 組のデータセットを作成し、8-fold Cross Validation を行った。主成分分析には、Raw 波形のみを用いた。

240 個を学習データとして用い、NTF により式 (3) の **A**, **B**, **C**, **T** を推定した。識別特徴量には、基底行列 **T** の各行ベクトルを低次元の特徴量として用いた。評価データには、学習データの分解時に得られる **A**, **B**, **C** をそのまま与え、**T** のみを推定することによって、得られた低次元ベクトルを用いた。このとき、次元数は $R = 20$ とし、式 (2) の左辺を基にし、右

辺の近似比率が 0.95 以上となるものを採用した。識別には、ガウシアンカーネルによる非線形マルチラス SVM を用いた。

4. 結果と考察

表 1 に加算回数が $N = 1$ (Raw 波形), 5, 10 (誘発波形) のときの識別の正答率を示す。全被験者の平均は、各波形 34.9 %, 36.4 %, 36.0 % となった。脳磁界データは、脳磁場そのものが微弱であるため、外因ノイズの影響を受けやすい。一般に、イベントに同期した脳反応 (誘発反応) を取りだすためには、イベントのオンセットに合わせた加算平均を行うが、加算回数が多いほど高い S/N 比の反応が得られる。表 1 の結果からは、被験者 3 を除くと、加算回数 $N = 5$ の時が最も正答率が高く、Raw 波形より誘発波形を用いたほうが正答率が高い傾向が見られた。一方、誘発波形同士の正答率の差はわずかであるものの、加算回数を増やすほど正答率が向上するような傾向があるとは言い難い。音声想起時の同期が低いために、加算回数の少ない場合にはノイズの除去が不十分となり、加算の組み合わせによっては却って学習及び評価に用いる波形が不安定なものとなった可能性がある。今回は 8-fold Cross Validation を行ったが、試行数を増やし、十分な信頼性を確保した上で議論する必要がある。

図 3 は、被験者 4 のデータを用いて計算された NTF の基底行列 A , B , C , T である。それぞれ、順に周波数成分、時間成分、チャンネル成分、試行成分および個々のテンソルの強度を表す。基底行列 A については、いずれの加算回数においても 1-13 Hz の範囲に高い値が見られ、とりわけ 1-4Hz に多くの成分が分解されているとわかる。基底行列 B については、誘発波形を用いた場合に想起の開始時と終了時に高い値をもつ基底が観測されており、加算平均を施すことで同期した脳活動の変化が大きく現れたと考えられる。基底行列 C については、上端下端が右の側頭部チャンネル・中央付近が左の側頭部チャンネルにおける成分を表す。チャンネル全体に高い値が分散しているものの、とりわけ中央周辺の基底に集中して観測されていることから、左側頭部周辺の脳活動成分が強いことがわかる。

また、全体として正答率は、チャンスレートよりやや高い値にとどまった。特徴量抽出に用いた NTF は、式 (3) をもとに多次元テンソルを分解し、周波数・時間・チャンネルを表すランク 1 テンソルの強度を表すと考えられるベクトルを特徴量に用いることで、低次元特徴ベクトルを生成することができたものの、やはり正答率は高いとは言い難い。識別に有効な情報は、ただ一つの基底ベクトルに含まれているわけではなく、他の基底と組み合わせた新しい特徴量の選択をする必要があるのかもしれない。

5. おわりに

想起時の脳磁界データからの特徴量識別における偏った結果の回避、および識別精度の向上を目的として、適切な次元削減処理を試みた。特徴量抽出に連続ウェーブレット変換と非負値テンソル分解を適用し、脳磁界データの多次元情報を損なわなような低次元な特徴量を用いた学習モデルの評価を行った。

しかしながら、全被験者の平均正答率は 35.7% であり、有効な次元削減効果が得られたとは言い難い。さらなる改善が必要である。

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。

文 献

- [1] R. Fazel-Rezai, B. Z. Allison, C. Guger, E. W. Sellers, S. C. Kleih and A. Kübler, "P300 brain computer interface : current challenges and emerging trends," *Frontiers in Neuroengineering*, vol.5, no.14, pp.1-15, July 2012.
- [2] L. Jäncke and N. J. Shah, "Hearing' syllables by 'seeing' visual stimuli," *European Journal of Neuroscience*, Vol.19, no.9, pp.2603-2608, May 2004.
- [3] M. Hoshiyama, A. Gunji and R. Kakigi, "Hearing the sound of silence: a magnetoencephalographic study," *NeuroReport*, Vol.12, no.6, pp.1097-1102, May 2001.
- [4] S. Uzawa, T. Takiguchi, Y. Arika and S. Nakagawa, "Spatiotemporal Properties of Magnetic Fields Induced by Auditory Speech Sound Imagery and Perception," *Proc. IEEE EMBC2017*, pp.2542-2545, Jeju, Korea, July 2017.
- [5] G. Pfurtscheller and F. H. Lopes da Silva, "Event-related EEG/MEG synchronization and desynchronization: basic principles," *Clinical Neurophysiology*, vol.110, no.11, pp.1842-1857, May 1999.
- [6] 宇澤 志保美, 滝口 哲也, 有木 康雄, 中川 誠司, "音声想起に伴う誘発脳磁界の時空間的特性", 第 31 回日本生体磁気学会大会論文集, vol.29, No.1, pp.104-105, June 2016
- [7] 矢野 肇, 滝口 哲也, 有木 康雄, 神谷 勝, 中川 誠司, "脳磁界計測を用いたエアコン音の聴感印象推定の試み -非負値テンソル分解による関連脳活動の抽出-", 日本音響学会 2016 年秋季研究発表会講演論文集, pp. 673-676, Sept. 2016.
- [8] A. Cichocki, R. Zdunek, A. H. Phan and S. Amari, "Non-negative Matrix and Tensor Factorizations: Applications to Exploratory Multi-Way Data Analysis and Blind Source Separation," Willey, 2009.