社団法人　電子情報通信学会
THE INSTITUTE OF ELECTRONICS,
INFORMATION AND COMMUNICATION ENGINEERS

信学技報
TECHNICAL REPORT OF IEICE.

[ポスター講演] 音想起に係る脳磁界反応の比較 ～ 等しいエンベロープをもつ音声と純音 ～

宇澤志保美[†,††]　滝口　哲也[†]　有木　康雄[†]　添田　喜治[††]　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, †††y.soeta@aist.go.jp,
††††s.nakagawa@chiba-u.ac.jp

あらまし　近年，脳活動を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，あらかじめ用意された選択肢の中から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．しかしながら，想起対象となる音に含まれる物理パラメータが脳反応与える影響については未だ不明な点が残されている．本研究では，想起した音声の振幅エンベロープと伴って生じる脳活動の関係を調査した．音声を想起時，および想起音声と等しい振幅エンベロープを持つ純音を想起時の脳磁界の時間-周波数特徴量を，連続ウェーブレット変換 (Continuous Wavelet Transform: CWT) を用いて推定した．主成分分析 (Principal Component Analysis : PCA) によって次元圧縮された特徴量を抽出し，サポートベクターマシン (Support Vector Machine: SVM) によって識別率を比較した．

キーワード　脳磁界計測，音声想起，振幅エンベロープ，純音

Investigation of Brain Magnetic Fields Associated with Sound Imagery ～ Speech and Pure Tone with Similar Envelopes ～

Shihomi UZAWA[†,††], Tetsuya TAKIGUCHI[†], Yasuo ARIKI[†], Yoshiharu SOETA[††], and Seiji NAKAGAWA[†††,††]

† Graduate School of System Infomatics, Kobe University 1–1 Rokkodai, Nada, Kobe, 657–8501, Japan
†† National Institute of Advanced Industrial Science and Technology (AIST) 1–8–31 Midorigaoka, Ikeda, 563–8577, Japan
††† Center for Frontier Medical Engineering, Chiba University 1–33 Yayoi, Inage, Chiba, 263–8522, Japan
E-mail: †uzawa@me.cs.scitec.kobe-u.ac.jp, ††{takigu,ariki}@kobe-u.ac.jp, †††y.soeta@aist.go.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, with most of the 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. However, it was unclear how each parameter of the imagined sound reflected to the brain response associated with sound imagery. Therefore, in this study, we investigated relationships between the amplitude-envelope of the imagined sound and brain activity. First, the time-frequency features were extracted from brain responses for the imagery of speech sounds and pure tones with same envelopes with speech sounds using a continuous wavelet transform (CWT) and a principal component analysis (PCA). Finally, the discriminative accuracy was compared using a support vector Machine (SVM).

Key words　magnetoencephalography (MEG), sound imagery, envelope of amplitude, pure tone

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，すなわち，ブレイン・コンピュータ・インターフェース (BCI) の開発が盛んに行われている．例えば，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タイプの 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行ってきた．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ところ，時間波形上で聴覚野周辺の緩やかな活動が認められたものの [4]，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．想起音声の違いを判別するためには，時間特徴量以外にも着目する必要があると思われたため，時間周波数特徴量に注目した [5]．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 [6]，音声想起でも同様な変動が観察される可能性は高い．微細な特徴量の変化を捉えるために機械学習を用いて解析を行ったが，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かった．音声想起の脳磁界データは，周波数や時間，さらにはチャネルなどの多次元情報をもつことから，非負値テンソル分解を用いた特徴量抽出による想起音声の識別も試みた [7]．全被験者の平均正答率はチャンスレートよりわずかに高い程度であったものの，振幅エンベロープの立ち上がり形状が他と異なる音声においては，高い精度が得られた．この結果は，想起音声のエンベロープの違いが識別精度に大きく影響する可能性を示している．

音声の想起には，意味の理解，過去の記憶との照合などの内的プロセスも含めた，様々な要素が影響すると思われる．しかしながら，想起対象となる音に含まれる各種の物理パラメータが，脳反応にどのように反映されるかという基礎的な点においてさえ不明な点が残されている．個々のパラメータがそれぞれに脳反応へ及ぼす影響の詳細を知ることによって，多数のパラメータを含有する複雑な脳反応の中から，着目すべき成分だけを抽出することが可能となる可能性がある．このことは，識別特徴量の低次元化にもつながると思われる．

本稿では，音のパラメータのうち，振幅エンベロープに着目し想起時の脳反応への影響を調べた．振幅エンベロープは音の強弱の時間変化に相当する特徴量であり，音声想起時の脳活動を利用した意思疎通の可能性を検討するためには，言語性に関わる特徴量が脳活動にどの程度影響するかについても調査する必要がある．したがって本研究では，音声を想起した際の脳活動と，想起音声と等しいエンベロープを持つ純音を想起した時の脳活動を比較することにより，振幅エンベロープの変化が脳磁界反応へ与える影響について検証した．

2. 計測方法

2.1 刺激音

音声刺激には，4 モーラの日本語単語音声，“あまぐも"，“いまふう"，および “えいゆう"を用いた．いずれも，親密度別音声データベース (FW03, NTT-AT) に含まれる女性話者音源 (fto) を利用した．純音刺激には，各音声刺激と等しいエンベロープをもつような純音を 3 種類用意した．

まず，図 1 に示すように各音声刺激の振幅に対して，各々適当なサンプル区間にわたって平滑化した音声信号のピークに沿うようにエンベロープを抽出した．抽出したエンベロープは，基本周波数 500 Hz，サンプリング周波数 48000 Hz の sin 波と掛け合わせた．

次に，作成された純音刺激および音声刺激に対して，A 特性音圧レベルに基づいた聴感印象の補正を行った．A 特性とは，人間の耳が周波数別に異なる感度を有することを考慮した周波数補正値のことである．これによって補正された音圧レベルは A 特性音圧レベルと呼ばれ，信号の音圧レベルを人間の聴感に近くなるように補正を行った音圧レベルのことである．使用し

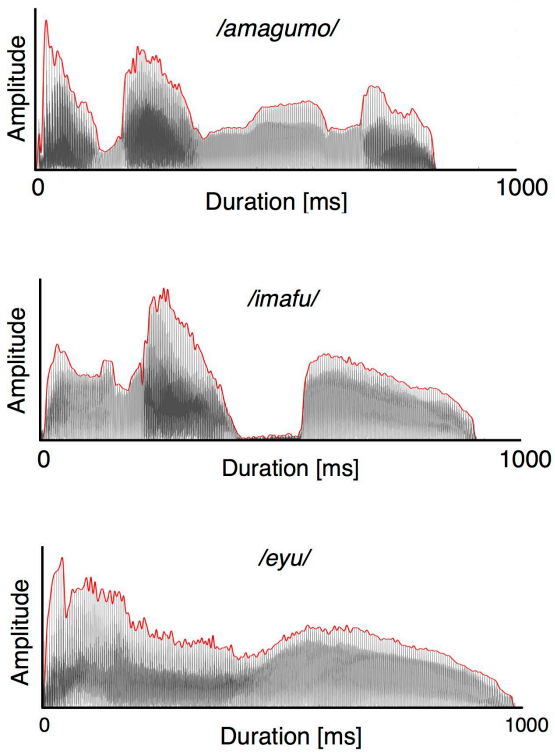

図 1　各音声の振幅エンベロープ
Fig. 1　The amplitude-envelopes of each speech sound.

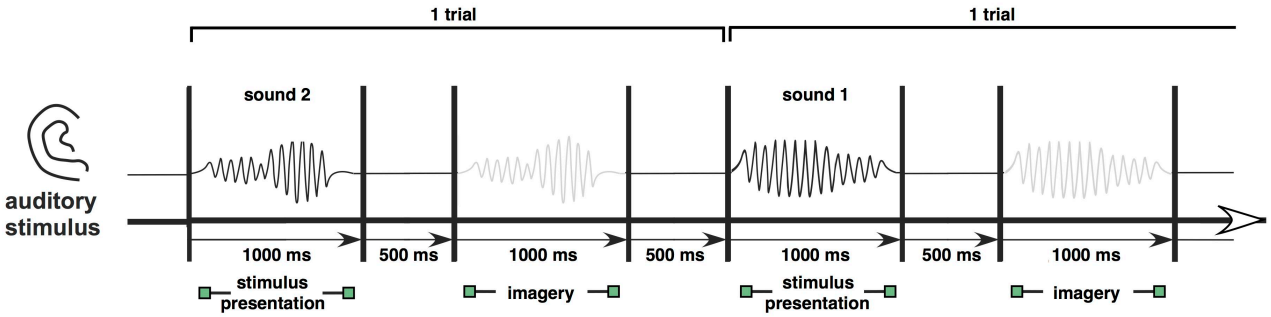

図 2　刺激呈示タスクの流れ
Fig. 2　Schematic diagram of the task.

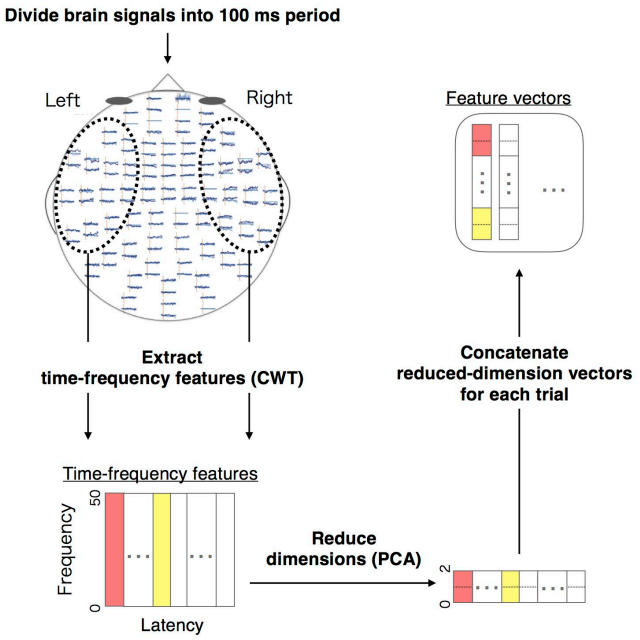

図 3　特徴量抽出の流れ
Fig. 3　Schematic diagram of the feature extraction.

たすべての刺激音は，A 特性下において，音声区間のパワーの 2 乗平均平方根が等しくなるように A 特性音圧レベルを修正した．加えて，計測を行う前には，被験者ごとに “はっきりと聞こえ，なおかつちょうど良いと感じる大きさ" かつ “音声刺激と純音刺激がちょうど同じように聞こえる大きさ"に設定した．

2.2　刺激呈示およびタスク

図 2 に刺激呈示の流れを示す．各試行では，持続時間 1000 ms の刺激音（単語音声）の後，500 ms の刺激間間隔 (Inter-stimulus interval : ISI) を挟んで，刺激音の持続時間と等しい無音区間が与えられる．

被験者には，無音区間において，直前に呈示された刺激音を想起するように求めた (想起条件)．ここで，本研究で用いる想起の定義は，口などは動かさず，第三者の声として呈示された音声をそっくりそのまま頭の中で強く再生することとする．また，被験者にタスクへの集中を持続させるために，一定確率で周波数の高い刺激音や異なる単語を呈示し，被験者にはその都度ボタンで応答させるようにした．同様の刺激呈示を行いながら，被験者に無音映画を見せることにより想起を全く行わせない条件下での計測も行った (無想起条件)．

被験者が純音刺激と音声刺激の関係を認識し，純音を想起する際に対応する音声を同時に想起してしまうことを防ぐため，全ての被験者に対して純音想起のタスク後に音声想起のタスクを実施した．また，想起条件のタスク後には無想起条件のタスクを挟み，合計 4 セッション行った．計測後にはアンケートを実施し，各被験者が純音刺激と音声刺激を如何に知覚したかを調べた．

刺激音の呈示開始時から 1500 ms 後を 0 ms とし，-100-1100 ms の区間における脳磁界データに対して 100 回以上加算平均を行った．

2.3　脳磁界計測

聴覚健常者 5 名 (男性 4 名，女性 1 名, 22-24 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 - 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

無音区間のうち，潜時-100~1100 ms を解析対象とした．解析には，想起区間の左右側頭部に相当する計 22 チャネルの脳磁界波形を用いて，純音想起波形と音声想起波形において想起音の識別精度に対する比較を行った．各被験者において，単語ごとに 100 試行分の脳磁界波形を用いており，そのうち，80 試行を学習データ，残りの 20 試行を評価データとした．

特徴量抽出の流れを図 3 に示す．得られた脳磁界波形を 100 ms ごとの潜時区間に分け，各区間において特徴量抽出および識別を行った．まず，以下の式に従って CWT [8] による時間-周波数特徴量の抽出を行った．

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

$f(t)$ は時系列脳磁界波形とする．$\psi(t)$ はウェーブレット関数

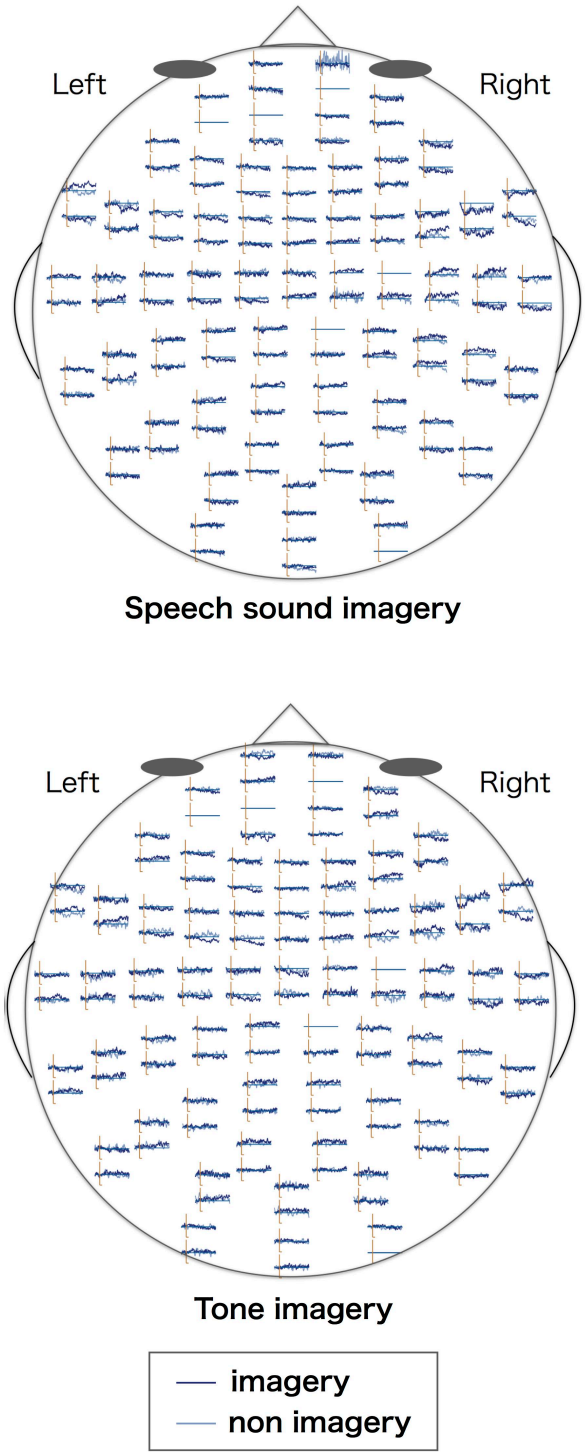

図 4　被験者 4 における音声想起時 (上図) と純音想起時 (下図) の誘発脳磁界の比較
Fig. 4　Brain evoked fields associated with speech sound imagery (top) and pure tone imagery (bottom) in subject 4.

(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．

次に，PCA による時間-周波数特徴量の次元削減を行った．CWT により得られた周波数 1-50 Hz の 50 次元時間-周波数特徴量について，フレーム，チャネルおよび試行回数分の周波数ベクトルを 50 次元から 2 次元へと削減した．削減後の特徴量に対して白色化を行った後，試行ごとに潜時区間の全フレーム (2.5ms ごとの 40 フレーム) と全チャネル (22 チャネル) の時間-周波数特徴量を連結し，識別特徴量を作成した．

学習および評価データについては，試行の組み合わせを入れ替えることによって 5 組のデータセットを作成し，5-fold Cross Validation を行った．識別には，ガウシアンカーネルによる非線形マルチクラス SVM を用いた．

4. 結　　果

音声想起時と純音想起時における想起および無想起条件について，図 4 に誘発反応波形 (周波数帯域 [0.1-30 Hz]) を示す．図 5 に全被験者の各セッションにおける識別結果を示す．また，表 1 には，全被験者の音声想起時と純音想起時の識別率の比較を示す．被験者 1，2 において，おおよそ音声想起より純音想起の識別率が高い結果が得られているものの，被験者 3，4，5 においては，両者には目立った差異は認められない．

図 5 に，それぞれ純音想起と音声想起時において高い識別率が得られた潜時帯を示す．純音想起においては，被験者 1，4 を除き潜時 0-200ms に高い識別率が得られている．一方，被験者によりバラツキが見られるものの，音声想起においては純音想起時に比べて潜時帯が遅い傾向にあった．また，想起条件のどちらにおいても上記のピークを超えると識別率は減少する傾向にあり，約 500-600ms 帯以降からベースライン（無想起条件）との差異がはっきりと認められなくなった．

5. 考　　察

5 名中 3 名の被験者では，純音想起時と音声想起時で識別率の有意な差異は認められなかった (図 5，表 1)．これは，音声と純音に共通する要素すなわち振幅エンベロープが識別に影響する主要な要素であるとすれば，この結果を説明できる可能性がある．また，残りの 2 名の被験者では，純音想起時には音声想起時に比べて高い識別率が得られた．この結果は，単純な音の想起の方が脳活動の特徴を捉えやすいことを示唆している．音声の想起には，前述のように，意味の理解や過去の記憶との照合などの内的プロセスも影響すると思われる．そのようなプロセスに対応する複数の活動源に基づく脳磁界反応が重なるため，より単純な音を想起した際の脳磁界波形の方が識別に必要な情報を抽出することが容易だった可能性も考えられる．また，純音想起時には潜時 0-200ms に高い識別率が得られており，打撃音を用いた先行研究 [3] における音想起時の脳反応の観測潜時と一致していることから，音想起時の脳反応と識別に関わる脳反応は重なって存在する可能性が考えられる．なお，今回は特徴量の抽出に CWT と PCA を用いたが，識別精度そのものが実用に耐えうるものとは言い難いことから，対象とする脳反応の識別に必ずしも有効であったとは言えない．音声の想起時のような高次脳反応から言語情報に基づいた識別を可能にするためには，クラスタリングによる次元圧縮や活動源推定結果と組み合わせた 3 次元的な特徴量を用いるなど，さらに特徴量の抽出方法を工夫した上で検討していく必要がある．

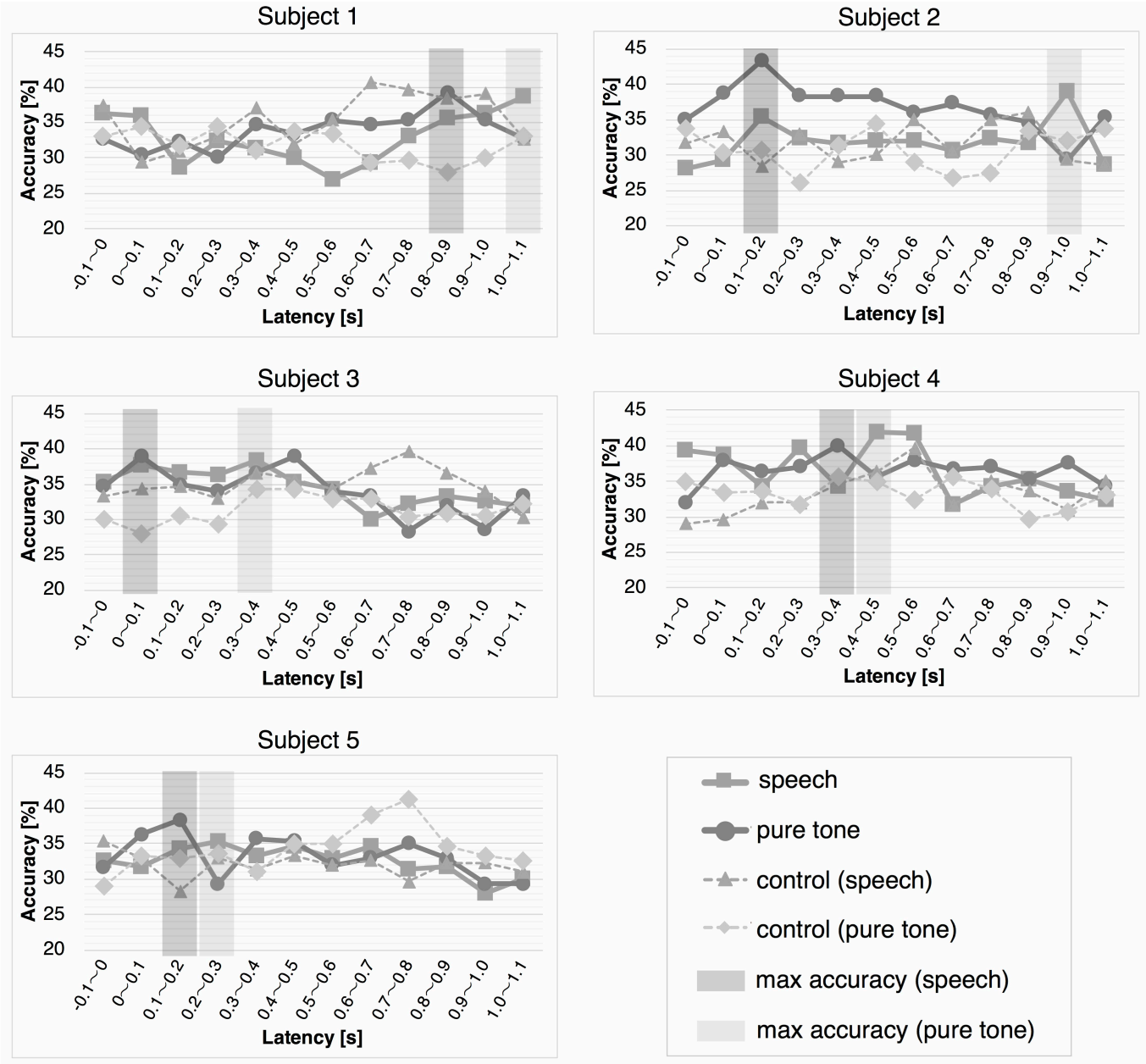

図 5　各被験者におけるセッションごとの想起単語の識別正答率

Fig. 5　Accuracy of the all sessions in each subject.

6. おわりに

本研究では，想起対象の音声の振幅エンベロープと脳活動の関係を調査した．まず，5 名の被験者に対して，3 種の音声刺激と，音声刺激と等しい振幅エンベロープを持つ純音刺激を想起した際の脳活動をそれぞれ計測した．計測した脳磁界を 100ms ごとの潜時区間に分割し，CWT による時間-周波数特徴量の抽出を行った．次に，PCA によって時間-周波数特徴量の次元削減を行い，試行ごとに特徴量ベクトルを作成した．最後に，SVM による想起単語の識別を行い，音声想起時と純音想起時の識別率を比較したところ，5 名中 3 名の被験者において，純音想起時と音声想起時で識別率の有意な差異は認められなかった．この結果から，刺激音の振幅エンベロープが識別に影響する主要な要素である可能性が示唆された．しかしながら，残る 2 名では，純音想起時には音声想起時に比べて高い識別率が得られており，今回の特徴量抽出では，処理が単純な音の想起時における識別の方が容易であった可能性が考えられる．そのため，音声想起時のように高次脳反応から言語情報に基づいた識別を可能にするためには，クラスタリングによる次元圧縮や神経電流源推定結果と組み合わせた 3 次元的な特徴量を用いるなど，より発展的な特徴量抽出法の検討が必要と思われる．

謝　　辞

本研究の一部は科学研究費補助金（17H02079, 25282053, 15H02771）の援助を受けて実施された．

表 1　各被験者における音声想起時と純音想起時の想起単語の識別正答率の比較

Table 1　Comparison with the discrimination accuracy of speech sound imagery and pure tone imagery.

	Speech sound imagery / Tone imagery					
Sub.	-100∼0 ms	0∼100 ms	100∼200 ms	200∼300 ms	300∼400 ms	400∼500 ms
1	36.3 / 32.7	36.0 / 30.3	28.7 / 32.3	32.3 / 30.0	31.3 / 34.7	30.0 / 33.3
2	28.0 / 35.0	29.3 / 38.7	35.3 / 43.3	32.3 / 38.3	31.7 / 38.3	32.0 / 38.3
3	35.3 / 34.7	37.7 / 39.0	36.7 / 35.0	36.3 / 34.0	38.3 / 36.7	35.3 / 39.0
4	39.3 / 32.0	38.7 / 38.0	34.3 / 36.3	39.7 / 37.0	34.3 / 40.0	42.0 / 35.7
5	32.7 / 31.7	31.7 / 36.3	34.3 / 38.3	35.3 / 29.3	33.3 / 35.7	34.7 / 35.3
Sub.	500∼600 ms	600∼700 ms	700∼800 ms	800∼900 ms	900∼1000 ms	1000∼1100 ms
1	27.0 / 35.3	29.3 / 34.7	33.0 / 35.3	35.7 / 39.3	36.3 / 35.3	38.7 / 32.7
2	32.0 / 36.0	30.7 / 37.3	32.3 / 35.7	31.7 / 34.7	39.0 / 29.3	28.7 / 35.3
3	34.3 / 34.0	30.0 / 33.3	32.3 / 28.3	33.3 / 32.0	32.7 / 28.7	32.0 / 33.3
4	41.7 / 38.0	31.7 / 36.7	34.3 / 37.0	35.3 / 35.3	33.7 / 37.7	32.3 / 34.3
5	33.0 / 32.0	34.7 / 33.0	31.3 / 35.0	31.7 / 33.3	28.0 / 29.3	30.0 / 29.3

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. Ariki 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] 宇澤志保美, 滝口哲也, 有木康雄, 中川誠司, “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-,” 日本音響学会 2016 年秋季研究発表会講演論文集, pp. 621-624, Sept. 2016.

[6] G. Pfurtscheller, F.H. Lopes da Silva, “Event-related EEG/MEG synchronization and desynchronization: basic principles” Clinical neurophysiology, vol. 110, pp. 1842-1857, 1999.

[7] 宇澤志保美, 滝口哲也, 有木康雄, 中川誠司, “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 年秋季研究発表会講演論文集, 3-P-8, pp. 337-340, Sept. 2017.

[8] C. Torrence, G. P. Compo, s“A practical Guide to Wavelet Analysis” Bulletin of the American Meteorological Society, vol. 79 (1), pp. 61-78, 1998.