

音想起に伴う脳磁界反応：等しいエンベロープをもつ音声と純音の比較*

◎宇澤志保美 (神戸大/産総研), 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

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, さらにチャン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. 音声想起時の脳活動と, 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 刺激呈示およびタスク

Fig. 1 に刺激呈示の流れを示す. 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

*Brain magnetic fields associated with sound imagery: Comparison between speech and pure tone with similar envelopes by UZAWA, Shihomi (Kobe Univ./AIST), TAKIGUCHI, Tetsuya, ARIKI, Yasuo (Kobe Univ.), SOETA, Yoshiharu (AIST), NAKAGAWA, Seiji (Chiba Univ./AIST).

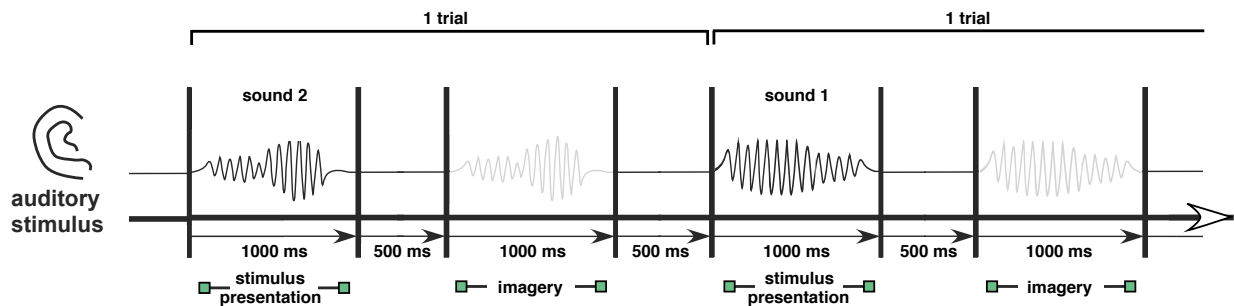


Fig. 1 Schematic diagram of the task.

刺激には、周波数 500Hz の純音に対して、各音声刺激から抽出したエンベロープを掛け合わせたものを用いた。それぞれ 3 種類の音声刺激、純音刺激が作成されたことになる。

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、直前に提示された刺激音を想起するように求めた（想起条件）。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ない条件下での計測も行った（無想起条件）。

計測では、純音想起タスクの後に音声想起タスクを行った。各想起条件タスク後には無想起条件タスクを挟み、合計 4 セッション行った。いずれのタスクにおいても各単語 100 回以上の計測を行った。また、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。

2.2 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™: 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 44 チャンネルの脳磁界波形を用いた。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。

連続ウェーブレット変換（Continuous Wavelet Transform: CWT）を用いて特徴量抽出を行い、主成分分析（Principal Component Analysis: PCA）を用いて次元削減を行った後、サポートベクタマシン（Support Vector Machine: SVM）を用いて想起音の識別を行った。特徴量の抽出および識別には、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、80 試行を学習データ、残りの 20 試行を評価データとした。

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。

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を 100 ms ごとの時間区間に分け、各区間におけるフレームとチャン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、試行ごとに時間区間の全フレーム（2.5ms ごとの 40 フレーム）と全チャンネル（44 チャンネル）の時間周波数特徴量を連結し、それらを特徴量ベクトルとした。

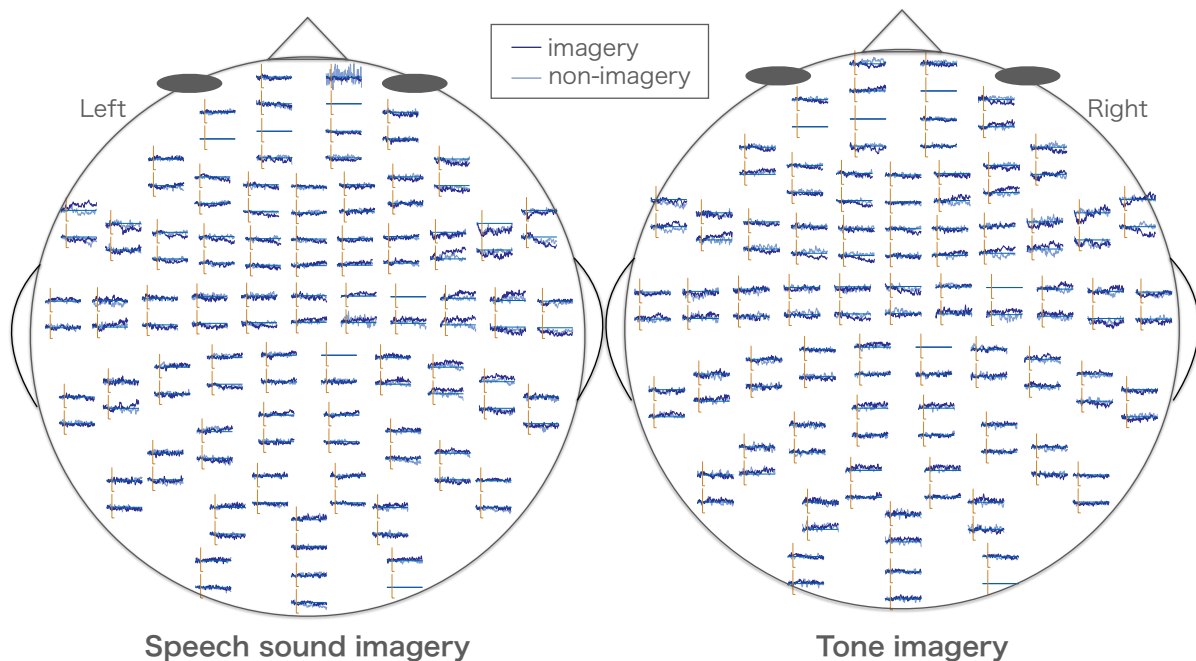


Fig. 2 Brain evoked fields associated with speech sound imagery(left) and pure tone imagery(right) in subject 4.

学習及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、Fig. 2に誘発反応波形(周波数帯域 [0.1-30 Hz])を、Fig. 3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、Table 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, 5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においては、両者には目立った差異は認められない。

Fig. 3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、純音想起時には音声想起時に比べて高い識別率が得られた (Fig. 3, Table 1)。この結果からは、単純な音の想起の方が脳活動としては捉えやすいことが示唆される。一方、残りの2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。刺激音の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かもしれない。加えて、前述のように、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だけを抽出することが容易だった可能性も考えられる。なお、今回は特徴量の抽出にCWTとPCAを用いたが、対象とする脳反応の識別に不向きだった可能性も否定はできない。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、特徴量の抽出方法を工夫した上での検討も必要であると考えられる。

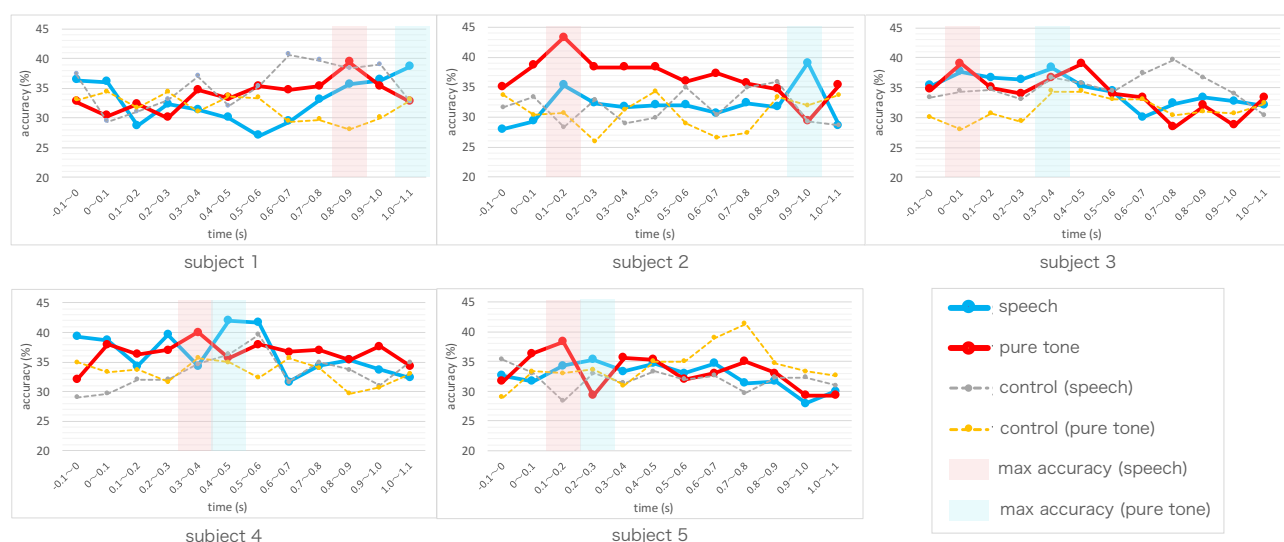


Fig. 3 Accuracy of the all sessions in each subject.

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

Sub.	Speech sound imagery / Tone imagery					
	-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

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。

参考文献

- [1] R. Fazel-Rezai *et al.*, Frontiers in Neuro-engineering, pp. 1-15, 2012.
- [2] L. Jäncke *et al.*, European Journal of Neuroscience, Vol. 19, pp. 2603-2608, 2004.
- [3] M. Hoshiyama *et al.*, NeuroReport, Vol. 12, pp. 1097-1102, 2001.
- [4] S. Uzawa *et al.*, In Proc. IEEE EMBC2017, pp. 2542-2545, 2017.
- [5] 宇澤ら, 日本音響学会 2016 年秋季研究発表会, 2016.
- [6] G. Pfurtscheller *et al.*, Clinical neurophysiology, vol. 110, pp. 1842-1857, 1999.
- [7] 宇澤ら, 日本音響学会 2017 年秋季研究発表会, 2017.