Decoding single-trial EEGs during silent Japanese words by the Transformer-like model

山﨑敏正[1] 赤迫健太[2] 森田寛伸[2] 徳永由布子[2]
上村旺生[2] 柳橋圭[2] 柏田倫孝[2]
[1]九州工業大学大学院情報工学研究院生命化学情報工学研究系 [2]九州工業大学情報工学部生命化学情報工学科
tymzk@bio.kyutech.ac.jp
{akasako.kenta289,morita.hironobu375,yuko.tokunaga124,uemura.ousei722,yanagibashi.kei127,kashiwada.mititaka602}@mail.kyutech.jp

概要

Silent speech interface 研究において、日本語単語を silent speech した時に頭皮上で記録される single-trial EEGs を解読して日本語で表現する方法を提案する．この方法は、日本語単語がすべて拍で表現可能である事、加算平均の考え方を参考に single-trial EEG から average ERPs を推定する方法を応用した事、Transformer の Attention 機能を活用した事、から構成される．その結果、5 個の相異なる拍の加算平均後の脳波を使って、サイレント「いえ（家）」と「いいえ」の single-trial EEG 解読の可能性を示せた．

1 はじめに

Silent speech interface（SSIs）とは、音声生成やそうしようとした時に生成される非音響的な生体信号からその音声を解読することにより、口頭会話を復元するための補助的手段である[1,2]．音声に関連した生体信号の様々なセンシングモダリティーの中で、音声生成に関連する脳領域における神経活動を直接的におよび間接的に捉えられるのが脳波である．

脳波を利用した silent-speech-to-text アプローチは、初めて Suppes, Lu & Hau[3]により 7 つのサイレント英単語で調べられ、後に音素[4,5]、音節[6]へ進んでいった．近年ではサイレント日本語 2-mora 単語をサイレントスピーチ（silent speech、SS）した時の頭皮脳波の解読[7]、2-mora 以上の単語を SS した時の脳波（加算平均後）の解読[8]が試みられている．より最近では、ヒアリングや発話時に記録された electrocorticography（ECoG）の解読が行われている[9,10,11]．

本研究では、[8]の発展として、日本語単語を SS した時の脳波（single-trial EEGs）の解読を試みる．本手法の neural network 構造は近年自然言語処理分野で多用されている Transformer [12]に類似している．

2 材料と方法

2.1 実験方法

被験者は 21 歳右利き男性学生ボランティア 3 名（被験者 A、B、C）である．尚、実験手順は「九州工業大学大学院情報工学研究院等における人を対象とする研究審査委員会」で承認されている．

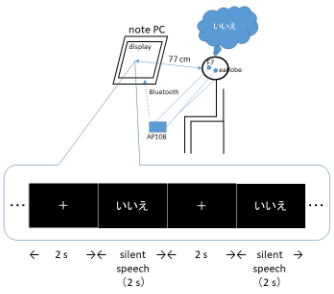

図1 本実験方法の全体像

最初に、被験者から 77 cm 前方に位置するディスプレイ上に注視点”＋”が 2s 間提示される．次に、SS すべき日本語単語がひらがなで 2s 間提示される．単語が提示されたら、被験者は出来る限り早く SS する．この計 4 秒間を 1 trial とする（図 1）．日本語は拍（mora）で数える言語であり[13]、mora は日本語単語生成の音韻的な単位である[14]．即ち、

日本語単語はすべてこの拍で表現することが出来る．図 2 は日本語拍すべてを示している[15]．同図の第 1 列目が日本語母音を示しており、最後の列“ん”、“っ”、“—”は、それぞれ、撥音、促音、長音と呼ばれ、最後の 2 拍は単独では発音が出来ない．

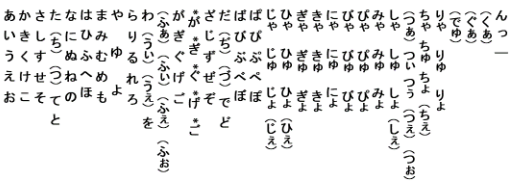

図2　日本語の拍．（）内の拍は外国語と感嘆詞．＊はが列の鼻音．

1 つのアクティブ電極 (AP-C151-015, ミユキ技研, 日本)が国際 10-20 システムに従い F7 に設置された．2 つの耳朶電極の平均を参照とした．これらの電極で記録された脳波はワイヤレス生体信号アンプ (Polymate Mini AP108, ミユキ技研, 日本) へ送られ、60 Hz の notch filter をかけ 10,000 倍に増幅された．サンプリング周波数は 500 Hz とした．エポック区間は単語提示の前後 2 s とした．オンラインで A/D 変換された脳波データは Bluetooth を介してすぐにパソコンに転送され、パソコン内のハードディスクに格納された (図 1 を見よ).

2.2　脳波データ解析方法

本研究の脳波解析モデルは以下のように encoder-decoder 型に類似している．加算平均の原理は single-trial EEG＝信号＋ノイズ（平均 0）を前提とする．このノイズ下で記憶すべきパターンを復元できる RNN（recurrent neural network）が知られている[16]（詳しくは付録参照）．上記信号および記憶すべきパターンを、各拍を SS した時の加算平均後の ERPs（event-related potentials）として RNN に入力すると、RNN の出力はノイズ下で各サイレント拍脳波の復元となる（拍が 106 個であれば 106 個の波形が得られる）．これが encoder の出力となる．

Decoder は single-trial EEG を入力として attention 機能が作用する．Scoring は encoder で得られた各サイレント拍脳波と single-trial EEG の内積として求まる．正規化された内積値から成る、拍数次元ベクトル $\boldsymbol{a}$ が得られる（各サイレント拍脳波がどの程度復元されるかを定量化する）．

更に、decoder への入力である single-trial EEG を時間的に重なり合うブロックに分割し、各ブロックに上記の処理を施せば、ブロック数分のベクトル $\boldsymbol{a}$ が得られる．尚、隣り合うブロック間の違いは 2 ms であり、ブロック数（N）は 50 とした．また、教師信号を取り込むために、Attention 層から context vector $\boldsymbol{c}$ を算出し、”Affine”と”Softmax with Loss”層を加えて backpropagation で学習させた．ここまでの本モデルの構造とデータ解析の流れを図 3 にまとめた．

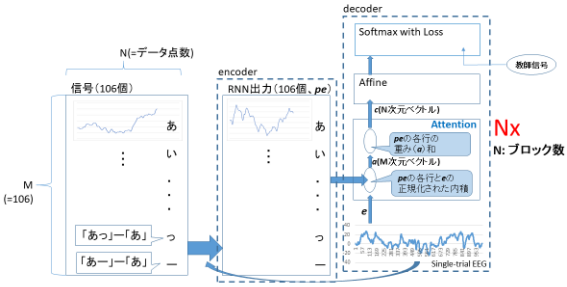

図3　本研究のneural network構造とデータ解析の流れ

3　結果

サイレント「いえ（家）」と「いいえ」の single-trial EEGs に本手法を適用した．結果をそれぞれ図 4 と図 5 に示した．

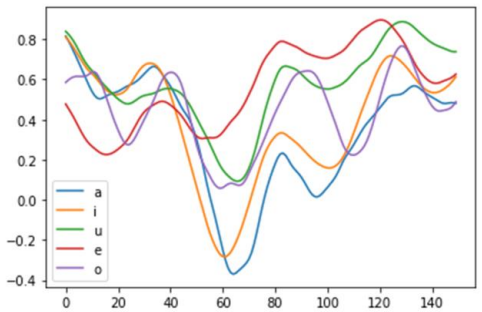

図4　Single-trial「いえ（家）」脳波を入力とした結果。横軸はブロック数、縦軸は正規化された内積値。

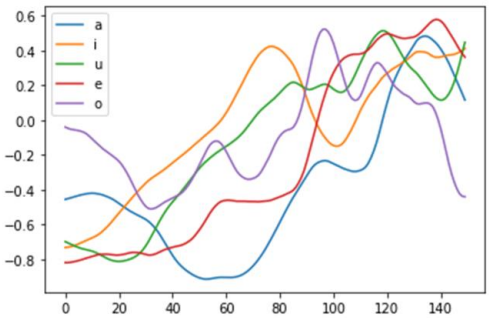

図5　Single-trial「いいえ」脳波を入力とした結果。横軸はブロック数、縦軸は正規化された内積値。

図 4 では、35 番目のブロック辺りで「い」の内積

値が最大となり、120 番目のブロック辺りで「え」の内積値が最大となった．図 5 では、90 番目のブロック辺りで「い」が最大となり、140 番目のブロック辺りで「え」の内積値が最大となった．ブロック数の違いに注目すれば、「いえ」と「いいえ」を検出して識別できると推察される。

4 おわりに

本研究の日本語単語を SS した時の single-trial EEG を解読する方法の特徴は、日本語単語がすべて拍で表現可能である事、加算平均の考え方を参考に single-trial EEG から average ERPs を推定する方法を適用した事、Transformer の Attention 機能を活用した事、である．今回例示した日本語単語が 2 つのみ、加算平均後の拍脳波が 5 つのみだったので、今後より多くの拍脳波を使って大多数の日本語単語で調べなければならない．また、解読のために vector ***a*** や context vector ***c*** をどう利用するか、本モデルの学習機能（図 3 の「Affine」と「Softmax with Loss」）の利用も今後の検討課題である．

参考文献

1. B. Denby, T. Schultz, K. Honda, T. Hueber, J. M. Gilbert, and J. S. Brumberg, Silent speech interfaces, Speech Commun., vol.52, no.4, pp.270-287, April 2010.
2. J. A. Gonzalez-Lopez, A. Gomez-Alanis, J. M. Martín Doñas, J. L. Pérez-Córdoba, and A. M. Gomez, Silent speech interfaces for speech restoration: A review, IEEE Access, vol.8, pp.177995-178021, 2020.
3. P. Suppes, Z.-L. Lu, and B. Han, Brain wave recognition of words, Proc. Natl. Acad. Sci., vol.94, pp.14965-14969, 1997..
4. C. S. DaSalla, H. Kambara, M. Sato, and Y. Koike, Single-trial classification of vowel speech imagery using common spatial patterns, Neural Networks, vol.22, pp.1334-1339, 2009.
5. M. Matsumoto, and J. Hori, Classification of silent speech using support vector machine and relevance vector machine, Applied Soft Computing, vol.20, pp.95-102, 2014.
6. M. D'Zmura, S. Deng, T. Lappas, S. Thorpe, and S. Srinivasan, Toward EEG sensing of imagined speech, in Human-Computer Interaction, Part I, HCII 2009, LNCS 5610, ed. J. A. Jacko, pp.40-48, Springer-Verlag Berlin Heidelberg.
7. H. Yamaguchi, T. Yamazaki, K. Yamamoto, S. Ueno, A. Yamaguchi, T. Ito, S. Hirose, K. Kamijo, H. Takayanagi, T. Yamanoi, and S.Fukuzumi, Decoding silent speech in Japanese from single trial EEGs: Preliminary results, Journal of Computer Science Systems Biology, vol.8, no.5, pp.285-292, 2015.
8. 山﨑敏正、徳永由布子、伊藤智恵子、Attention-based RNN with encoder-decoder によるサイレント日本語単語脳波の解読、電子情報通信学会　HCS 研究会、2023 年 1 月 22 日（発表予定）.
9. C. Herff, D. Heger, A. de Pesters, D. Telaar, P. Brunner, G. Schalk, and T. Schultz, Brain-to-text: decoding spoken phrases from phone representation in the brain, Frontier in Neuroscience, vol.9, pp.217-, 2015.
10. D. A. Moses, M. K. Leonard, J. G. Makin, E. F. Chang, Real-time decoding of question-and-answer speech dialogue using human cortical activity, Nature Communication, vol.10, pp.3096-, 2019.
11. J. G. Makin, D. A. Moses, and E. F. Chang, Machine translation of cortical activity to text with an encoder-decoder framework, Nature Neuroscience, vol.23, pp.575-582, 2020.
12. A. Vaswani, N. Shazeer, N. Parmer, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, Attetion is all you need, arXiv preprint arXiv: 1706.03762v5, 6 Dec 2017.
13. N. S. Trubetzkoy, Grundzüge der Phonologie (Principles of Phonology), Los Angeles: University of California Press, 1958/69.
14. B. G. Verdonschot, S. Tokimoto, and Y. Miyaoka, The fundamental phonological unit Japanese word production: An EEG study using the picture-word interference paradigm, Journal of Neurolinguistics, vol.51, pp.184-193, 2019.
15. 金田一春彦、日本語　新版（上）、岩波書店、第 53 刷、2017.
16. R. Laje, and D. V. Buonomano, Robust timing and motor patterns by taming chaos in recurrent networks, Nat. Neurosci., vol.16, no.7, pp.925-933, 2013.
17. H. Lee, and S. Choi, PCA+HMM+SVM for EEG pattern classification. Proceedings of Seventh International Symposium on Signal Processing and Its Applications, 2003.
18. B. Khalighinejad, G. C. da Silva, and N. Mesgarani, Dynamic encoding of acoustic features in neural responses to continuous speech, Journal of Neuroscience, vol.37, no.8, pp.2176-2185, 2017.
19. R. C. Pratap, The speech evoked potential in normal subjects and patients with cerebral hemispheric lesions, Clinical Neurology and Neurosurgery, vol.89, no.9, pp.237-242, 1987.
20. S. Tsukiyama, and T. Yamazaki, Discriminability among Japanese vowels using early components in silent-speech-related potentials, IEICE Technical Report, SP2019-31, WIT2019-30 (2019-10), pp.81-86, 2019.
21. M.-T. Luong, H. Pham, and C. D. Manning, Effective approaches to attention-based neural machine translation, ArXiv preprint arXiv: 1508.04025, 2015.
22. K. Cho, B. van Merriënboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, and Y. Bengio, Learning phrase representations using RNN encoder-decoder for statistical machine translation, arXiv preprint arXiv: 1406.1078, 2014.
23. I. Sutskever, O. Vinyals, and Q. V. V. Le, "Sequence to sequence learning with neural networks," Advances in Neural information Processing Systems, vol.27, pp.3104-3112, 2014.
24. G. K. Anumanchipalli, J. Chartier, and E. F. Chang, Speech synthesis from neural decoding of spoken sentences,

Nature, vol.568, pp.493-498, 2019.

25. L. Pratt, J. Mostow, and C. Kamm, “Direct transfer of learned information among neural networks,” Proc. the Ninth National Conference on Artificial Intelligence, vol.2, pp.584-589, AAAI Press, 1991.

付録

本研究で利用した RNN[16]の詳細は下図の通りである.

$$\tau \frac{dx_i}{dt} = -x_i + \sum_{j=1}^{N} W_{ij}^{Rec} r_j + \sum_{j=1}^{2} W_{ij}^{In} y_j + I_i^{noise} \quad (1)$$

r_i=tanh(x_i), y_j(j=1,2):入力ユニット

$$z = \sum_{j=1}^{N} W_j^{Out} r_j$$

Single-trial EEG = 信号＋ノイズ であるから、
ノイズ ＝「single-trial EEG ー信号」と推定

$$W_{ij}^{Rec}(t) = W_{ij}^{Rec}(t - \Delta t) - e_i(t) \sum_{k \in B(i)} P_{jk}^{i}(t) r_k(t)$$

$$e_i(t) = r_i(t) - R_i(t) \qquad R_i: \text{innate pattern}$$

$$P_{jk}^{i}(t) = P_{jk}^{i}(t - \Delta t) - \frac{\sum_{m \in B(i)} \sum_{n \in B(i)} P_{jm}^{i}(t - \Delta t) r_m(t) r_n(t) P_{nk}^{i}(t - \Delta t)}{1 + \sum_{m \in B(i)} \sum_{n \in B(i)} r_m(t) P_{mn}^{i}(t - \Delta t) r_n(t)}.$$

B(i): ユニットiに関して前シナプス的なrecurrentユニットの部分集合

図 ノイズ下で記憶すべきパターン(innate pattern)を復元できるRNN[16]