SAND2020-13976C

Visual Decoding of Phrases from Occipital Neuromagnetic Signals

Debadatta Dash[1], Paul Ferrari [2], Amir Borna[3], Joonas Iivanainen[3], Peter D. D. Schwindt[3], and Jun Wang[4]

Abstract— **Orthographic visual perception (reading) is encoded via a widespread dynamic interaction between different language centers of the brain and visual cortex. In this study, we investigated orthographic visual perception decoding with Magnetoencephalography (MEG), where phrases were visually presented to participants. We compared the decoding performance obtained with sensors within the occipital lobe that with whole-head sensors. Two naive machine learning classifiers namely support vector machines (SVM) and linear discriminant analysis (LDA) were used. Experimental results indicated that the decoding performance using only occipital sensors is similar to the performance obtained with all sensors within the task period, which were all above chance level. In addition, temporal analysis by taking short-time windows showed that the occipital sensors were more discriminative near onset compared to later time periods, while using the whole head sensor setup at later time periods performed slightly better than occipital sensors. This finding may indicate a sequential order (from vistual cortex to other areas beyond occipital lobe) during visual language perception.**

I. INTRODUCTION

Visual perception is a neural mechanism that allows the brain to create patterns of activity to receive, interpret, and act upon visual stimuli through a series of transformations of neural signals [1]. Orthographic visual perception is a "mid-level vision" process that acts as a central interface between visual and linguistic processing during reading [2]. The role of the occipital cortex, specifically ventral occipital-temporal cortex (vOT) is found to be crucial by several functional magnetic resonance imaging (fMRI) studies [3]–[6]. In regards to recognition of orthographic perception, several studies have attempted to explore pattern analysis using electroencephalography (EEG) [7], [8], fMRI [9], [10], electrocorticography (ECoG) [11], and magnetoencephalography (MEG) [12]. Very recently, beyond classification, research studies have attempted even for neural visual image reconstruction [13], [14]. These studies have focused on word-based stimu;i which constrains the complete understanding of visual lexico-semantic information that can be generated with a complete sentence/phrase. In addition, we hypothesize that the neural decoding paradigm on short-time windows might suggest a sequential order of visual language perception, where the neural activities start from visual cortex and then progress to other areas.

In this study, we used phrases as stimuli and explored the decoding of visual perception of language and the role of the occipital lobe as a measure of decoding performance. We used MEG to record the neural activity of healthy subjects while visually perceiving different phrases. MEG is a non-invasive neuroimaging modality which measures post-synaptic neuronal current-induced magnetic fields using highly sensitive magnetometers and gradiometers. It has a good spatial resolution (2–3 mm) and an excellent temporal resolution (< 1 ms) which make this modality ideal to study fast and dynamic processes such as language perception. The magnetic permeability of biological tissues (dura, scalp, skull) is similar to that of empty space and so the magnetic field recorded by MEG remain undistorted, which is a great advantage over EEG [15]. Moreover, MEG has been proven effective for numerous temporal dependence of visual pattern analysis research including optimal image discrimination [16], exploration on time-varying representation of visual patterns [17], decoding stimulus information while manipulating visual consciousness [18], time-series decoding of object recognition [19]. The high temporal resolution of MEG makes it advantageous over fMRI to explore on the temporal dynamics of visual perception.

The objective of this study was to compare the decoding performance of occipital sensors with the whole-head sensor setup to understand the role of the visual cortex in orthographic visual perception. Two machine learning classifiers, support vector machines (SVM) and linear discriminant analysis (LDA), were used to classify 5 different phrase stimuli from the MEG recordings. To further understand, the temporal dependency of perception mechanism, we compared the decoding accuracies from the occipital sensors with whole-head sensors by taking different short-time windows. Overall, the results show that the occipital sensors can decode the phrases well above chance level similar to while decoding with all sensors.

II. MEG METHODS

We used two identical Triux Neuromag MEG devices (MEGIN, LCC) (Figure 1) to collect neuromagnetic signals

This work was supported by the University of Texas System Brain Initiative under award number 362221 and partly by the National Institutes of Health (NIH) under award numbers R03DC013990 and R01DC016621.

[1]Debadatta Dash is a graduate research assistant in the Department of Neurology, Dell Medical School and Ph.D. candidate in the Department of Electrical and Computer Engineering at The University of Texas at Austin, TX 78712, USA debadatta.dash@utexas.edu

[2]Paul Ferrari is the MEG Technical Director at Spectrum Health and Helen Devos Children's Hospital MI 49503, USA paul.ferrari@spectrumhealth.org

[3]Amir Borna is with the Physics Based Microsystems Division at the Sandia National Laboratories, Albuquerque, NM 87123 aborna@sandia.gov

[3]Joonas Iivanainen is a postdoc researcher with the Physics Based Microsystems Division at the Sandia National Laboratories, Albuquerque, NM 87123 jaiivan@sandia.gov

[4]Jun Wang is an Associate Professor in the Department of Speech, Language, and Hearing Sciences and in the Department of Neurology, Dell Medical School at the University of Texas at Austin, TX 78712, USA jun.wang@austin.utexas.edu

Fig. 1. The MEG unit with a subject and a stimulus “I need help” displayed on the screen

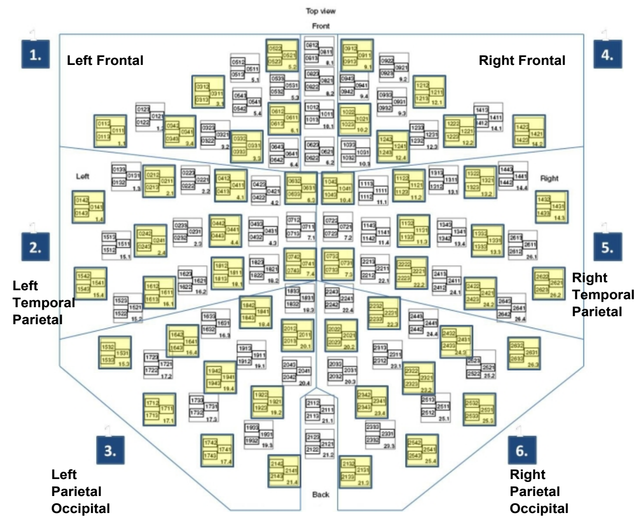


Fig. 2. The MEG sensor map: Modified with permission, from user manual ”TriuxTM” neo Instruction for Use, 2020 (MEGIN, LCC)

from 6 healthy subjects (2 females) with informed consent. We collected data at Dell Children’s Medical Center, Austin, TX, USA and Cook Children’s Medical Center, Fort Worth, TX, USA in compliance with the institutional review boards (IRB) of the participating institutions. The MEG system has 306 channels with 102 magnetometer sensors and 204 gradiometer sensors. The machine is housed inside a magnetically shielded room (MSR) to restrict external magnetic fields. we used 5 visual stimuli such as: *1. Do you understand me; 2. That’s perfect; 3. How are you; 4. Good-bye; 5. Ineed help*, those were displayed on a screen, one at a time, written in English. A stimulus dedicated computer running the STIM2 software (Compumedics, LTD) connected to a high-quality DLP projector was used to display the stimuli onto a back-projection screen situated at 90 cm from the subjects. Each stimulus was 1 s following a pre-stmuli interval of 0.5 s for 100 repetitions/trials in a pseudo-randomized order.

Data were recorded with a 4 kHz sampling rate with an online hardware filter of 0.03–1000 Hz. Bipolar electrooculography (EOG) and electrocardiograph (ECG) sensors were used for recording the eye movement and cardiac signals, respectively. The MEG signals were low-pass filtered below 250 Hz with a 4^{th} order Butterworth filter and resampled to 1 kHz. Line noise (60 Hz) and harmonics were removed with a notch filter. Only gradiometer sensor data were considered for decoding due to their effectiveness in noise suppression. Sensors that showed a flat or overly noisy response were discarded from analysis. Through visual inspection, trials containing large artifacts were removed with an average of 25% rejection rate. Data for one subject had only 63 valid trials for a phrase after preprocessing. Thus, for an unbiased comparison, we considered only the first 60 trials per phrase per subject for decoding. After preprocessing the signals were further processed through a db-4 wavelet denoising with 2 levels. Root mean square (RMS) values of the signals were computed as features for training the decoders. RMS features have been proven to be effective in language-based decoding studies [20]–[24].

III. Decoding Methods

A. Classifiers

We used two standard machine learning classifiers as decoders for the study. They are: support vector machines (SVM) and linear discriminant analysis (LDA). We used a second order polynomial kernel SVM following our previous studies [25]–[27]. The hyperparameters for SVM such as Kernel scale and C parameter were tuned and selected based on Bayesian optimization search. Automatic hyperparameter optimization was performed to find the best LDA parameters for Dirichlet distributions. The classification was performed for each subject with a 5-fold cross-validation (CV) strategy. For bootstrapping, the 5-fold CV was performed 10 times. The average accuracy across the 5 folds and 10 bootstrap runs was taken as the final performance.

B. Experiments

We performed two experiments to investigate and compare the visual decoding of phrases from occipital sensors and all (whole-head) sensors. First, we trained the decoders with the RMS features extracted from all sensors (~ 200) and then trained the decoders with RMS features extracted from occipital sensors only (~ 60) (Figure 2). Sections 3 and 6 of The MEG sensor map shown in Figure 2 which represent left and right parietal occipital cortex respectively were chosen as occipital cortex sensors altogether. Finally, we performed a temporal analysis where we extracted the RMS features from the perception segments with an increasing window of 100 ms starting from the stimulus onset and also with a decreasing time period of 100 ms from the onset. In other words, for increasing window analysis, the 10 time segments were [$1 - 100\ ms$; $1 - 200\ ms$; $1 - 300\ ms, ...1 - 1000\ ms$]. Similarly for decreasing window analysis, the time segments were [$1 - 1000\, ms$; $101 - 1000\, ms$; $...901 - 1000\, ms$]. We computed RMS features and performed decoding from each of these time segments.

IV. Results and Discussions

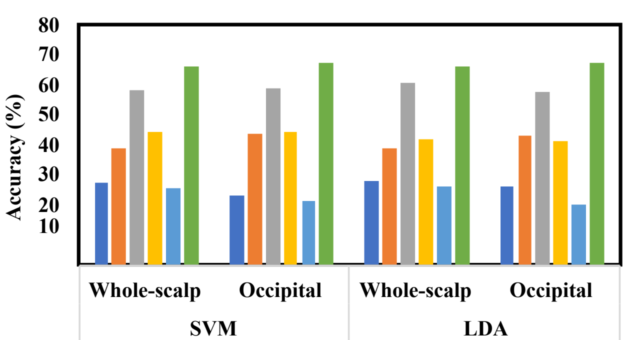
A. Occipital Lobe vs Whole Scalp

Figure 3 shows the comparison of decoding accuracies of visual phrases using all sensors and occipital sensors only for each subject using 2 classifiers. It shows that a similar accuracy was obtained with only occipital sensors (average=42.65%±3.01%) when compared to all sensors (average=43.5%±2.74%). This indicates that only occipital sensors might be sufficient for visual perception decoding. However, it might be also due to the possible inclusion of Wernicke's area sensors in the left parietal occipital cortex that is contributing to a similar accuracy. The performance of LDA and SVM was very similar as expected. Another important aspect to notice here is the consistency of performance for different subjects for these two groups (All and occipital). For example, the decoding accuracy was highest for Subject 6 while taking all sensors and also while taking occipital sensors irrespective of the decoders. Similarly, the performance of subject 5 was the lowest irrespective of the decoder and sensor group under consideration. Please note that the accuracies obtained for each subject and the average accuracy was higher than chance level (20% for a 5-class class classification).

These findings suggested that we may not need the whole-scalp sensors (~200) in MEG for visual perception decoding. Recently developed optically pumped magnetometer (OPM)-MEG [28]–[30] has shown potential for a wearable, low-cost, higher signal-to-noise ratio, and even shielding-free MEG. Our findings provide a support to reduce the number of sensors in OPM-MEG but can still obtain the similar performace in visual speech perception, which will significantly reduce the cost of the system.

B. Temporal Analysis

The results for temporal analysis averaged across subjects is shown in Figure 4 where it can be seen that for individual chunks of time occipital sensors perform slightly lower than



0

when decoding was performed using all sensors. Interestingly, when the initial temporal information was removed the decoding accuracy was decreased with time for occipital sensors but remained more or less the same when all sensors were used as can be seen from decreasing window analysis. This suggests that the occipital sensors might be contributing to the pattern recognition during initial periods but for visual perception, sensors beyond occipital cortex might be contributing.

A 1-tail t-test was conducted between the performances of the short-time windows which resulted that the decoding performance of the whole-scalp setup is significantly higher than occipital lobe sensors only ($p < 0.05$) for both increasing and decreasing window analysis. These findings may suggest a sequential order of visual language perception, where the neural activities start from visual cortex and then to other areas. However, the difference in performance is just in order of 3-4%, thus, thus, further studies with larger number of participants are need to verify these findings.

V. Conclusions

In this study, we investigated the decoding of visual perception of phrases using MEG. We used two standard machine learning classifiers and compared the decoding performance of occipital sensors with all sensors and found that they perform similarly for decoding. MRMR based feature ranking method indicated that a higher number of sensors have better predictor score in whole-head sensors compared to occipital sensors only. Windowed temporal analysis indicated that the similar performance of occipital sensors and all sensors were mostly during the initial identification stage but the performance of occipital sensors decreased in the later time segments. Future study will focus on reconstruction of phrases from the neuromagnetic activity.

VI. Acknowledgment

We thank Drs. Angel W. Hernandez-Mulero and Saleem Malik for their help on the data collection. We also thank Dr. Ted Mau, Dr. Myungjong Kim, Dr. Mark McManis, Dr. Daragh Heitzman, Kristin Teplansky, Saara Raja, and the volunteering participants. Sandia National Laboratories is a multimission laboratory managed and operated by National Technology & Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International Inc., for the U.S. Department of Energy National Nuclear Security Administration under contract DENA0003525. This paper describes objective technical results and analysis. Any subjective views or opinions that might be expressed in the paper do not necessarily represent the views of the U.S. Department of Energy or the United States Government.

References

[1] O. Braddick, "Visual perception, neural basis of," pp. 16269 – 16274, 2001.

[2] Jonathan Grainger, "Orthographic processing: A mid-levelvision of reading: The 44th sir frederic bartlett lecture," *Quarterly Journal of Experimental Psychology*, vol. 71, no. 2, pp. 335–359, 2018.

[3] Stanislas Dehaene and Laurent Cohen, "The unique role of the visual word form area in reading," *Trends in cognitive sciences*, vol. 15, no. 6, pp. 254–262, 2011.

Sub1 **Sub2** **Sub3** **Sub4** **Sub5** **Sub6**

Fig. 3. Comparison of decoding accuracies with All and occipital sensors for each subject

[4] Laurent Cohen, Stanislas Dehaene, Lionel Naccache, Stephane Lehéricy, Ghislaine Dehaene-Lambertz, Marie-Anne Hénaff, and François Michel, "The visual word form area: spatial and temporal characterization of an initial stage of reading in normal subjects and posterior split-brain patients," *Brain*, vol. 123, no. 2, pp. 291–307, 2000.

[5] Cathy J Price and Joseph T Devlin, "The interactive account of ventral occipitotemporal contributions to reading," *Trends in cognitive sciences*, vol. 15, no. 6, pp. 246–253, 2011.

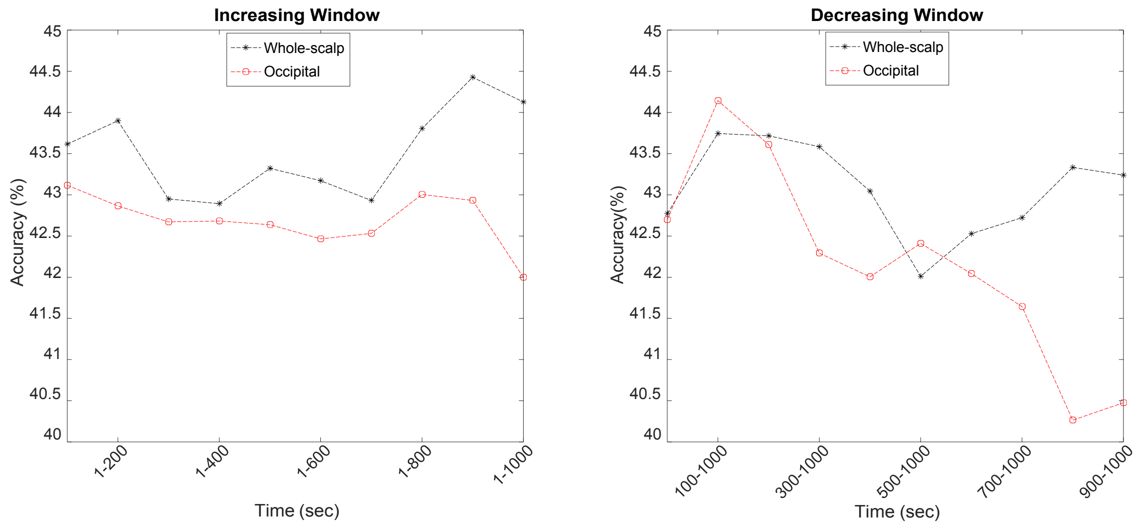


Fig. 4. Comparison of temporal analysis

[6] JSH Taylor, Kathleen Rastle, and Matthew H Davis, "Can cognitive models explain brain activation during word and pseudoword reading? a meta-analysis of 36 neuroimaging studies.," *Psychological bulletin*, vol. 139, no. 4, pp. 766, 2013.

[7] Patrick Suppes, Zhong-Lin Lu, and Bing Han, "Brain wave recognition of words," *Proceedings of the National Academy of Sciences*, vol. 94, no. 26, pp. 14965–14969, 1997.

[8] Shouyu Ling, Andy C. H. Lee, Blair C. Armstrong, and Adrian Nestor, "How are visual words represented? insights from EEG-based visual word decoding, feature derivation and image reconstruction," *Human Brain Mapping*, vol. 40, no. 17, pp. 5056–5068, 2019.

[9] Annelies Baeck, Dwight Kravitz, Chris Baker, and Hans P Op de Beeck, "Influence of lexical status and orthographic similarity on the multi-voxel response of the visual word form area," *Neuroimage*, vol. 111, pp. 321–328, 2015.

[10] Adrian Nestor, Marlene Behrmann, and David C Plaut, "The neural basis of visual word form processing: a multivariate investigation," *Cerebral Cortex*, vol. 23, no. 7, pp. 1673–1684, 2013.

[11] Elizabeth A Hirshorn, Yuanning Li, Michael J Ward, R Mark Richardson, Julie A Fiez, and Avniel Singh Ghuman, "Decoding and disrupting left midfusiform gyrus activity during word reading," *Proceedings of the National Academy of Sciences*, vol. 113, no. 29, pp. 8162–8167, 2016.

[12] Alexander M Chan, Eric Halgren, Ksenija Marinkovic, and Sydney S Cash, "Decoding word and category-specific spatiotemporal representations from MEG and eeg," *Neuroimage*, vol. 54, no. 4, pp. 3028–3039, 2011.

[13] Le Chang and Doris Y Tsao, "The code for facial identity in the primate brain," *Cell*, vol. 169, no. 6, pp. 1013–1028, 2017.

[14] Guohua Shen, Tomoyasu Horikawa, Kei Majima, and Yukiyasu Kamitani, "Deep image reconstruction from human brain activity," *PLoS computational biology*, vol. 15, no. 1, pp. e1006633, 2019.

[15] Sanjay P Singh, "Magnetoencephalography: basic principles," *Annals of Indian Academy of Neurology*, vol. 17, no. Suppl 1, pp. S107, 2014.

[16] ME Van de Nieuwenhuijzen, AR Backus, Ali Bahramisharif, Christian F Doeller, Ole Jensen, and Marcel AJ van Gerven, "MEG-based decoding of the spatiotemporal dynamics of visual category perception," *Neuroimage*, vol. 83, pp. 1063–1073, 2013.

[17] Susan G Wardle, Nikolaus Kriegeskorte, Tijl Grootswagers, Seyed-Mahdi Khaligh-Razavi, and Thomas A Carlson, "Perceptual similarity of visual patterns predicts dynamic neural activation patterns measured with MEG," *Neuroimage*, vol. 132, pp. 59–70, 2016.

[18] Anh-Thu Mai, Tijl Grootswagers, and Thomas A Carlson, "In search of consciousness: Examining the temporal dynamics of conscious visual perception using MEG time-series data," *Neuropsychologia*, vol. 129, pp. 310–317, 2019.

[19] Erika W Contini, Susan G Wardle, and Thomas A Carlson, "Decoding the time-course of object recognition in the human brain: From visual features to categorical decisions," *Neuropsychologia*, vol. 105, pp. 165–176, 2017.

[20] Lei Wang, Ed X. Wu, and Fei Chen, "Contribution of RMS-Level-Based Speech Segments to Target Speech Decoding Under Noisy Conditions," in *Proc. Interspeech 2020*, 2020, pp. 121–124.

[21] Debadatta Dash, Paul Ferrari, Saleem Malik, and Jun Wang, "Overt speech retrieval from neuromagnetic signals using wavelets and artificial neural networks," in *2018 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*. IEEE, 2018, pp. 489–493.

[22] Alborz Rezazadeh Sereshkeh, Robert Trott, Aurélien Bricout, and Tom Chau, "EEG classification of covert speech using regularized neural networks," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 25, no. 12, pp. 2292–2300, 2017.

[23] Debadatta Dash, Alan Wisler, Paul Ferrari, and Jun Wang, "Towards a speaker independent speech-BCI using speaker adaptation.," in *INTERSPEECH*, 2019, pp. 864–868.

[24] Debadatta Dash, Paul Ferrari, and Jun Wang, "Decoding imagined and spoken phrases from non-invasive neural (MEG) signals," *Frontiers in Neuroscience*, 2020.

[25] Debadatta Dash, Alan Wisler, Paul Ferrari, Elizabeth Moody Davenport, Joseph Maldjian, and Jun Wang, "MEG sensor selection for neural speech decoding," *IEEE Access*, vol. 8, pp. 182320–182337, 2020.

[26] Debadatta Dash, Paul Ferrari, Satwik Dutta, and Jun Wang, "NeuroVAD: Real-time voice activity detection from non-invasive neuromagnetic signals," *Sensors*, vol. 20, no. 8, pp. 2248, 2020.

[27] Debadatta Dash, Paul Ferrari, and Jun Wang, "Role of brainwaves in neural speech decoding," in *Proc. IEEE 28th Eur. Signal Process. Conf.*, 2020, pp. 1–5.

[28] Elena Boto, Niall Holmes, James Leggett, Gillian Roberts, Vishal Shah, Sofie S Meyer, Leonardo Duque Muñoz, Karen J Mullinger, Tim M Tierney, Sven Bestmann, et al., "Moving magnetoencephalography towards real-world applications with a wearable system," *Nature*, vol. 555, no. 7698, pp. 657–661, 2018.

[29] Amir Borna, Tony R Carter, Anthony P Colombo, Yuan-Yu Jau, Jim McKay, Michael Weisend, Samu Taulu, Julia M Stephen, and Peter DD Schwindt, "Non-invasive functional-brain-imaging with an opm-based magnetoencephalography system," *Plos one*, vol. 15, no. 1, pp. e0227684, 2020.

[30] Xin Zhang, Chun-qiao Chen, Ming-kang Zhang, Chang-yu Ma, Yin Zhang, Hui Wang, Qing-qian Guo, Tao Hu, Zhao-bang Liu, Yan Chang, et al., "Detection and analysis of MEG signals in occipital region with double-channel opm sensors," *Journal of Neuroscience Methods*, vol. 346, pp. 108948, 2020.