Decoding Semantic Word Categories from Electro- and Magnetoencephalography data

Simeon Spasov(1), Olaf Hauk(2), Seyedeh-Rezvan Farahibozorg (3)

(1) University of Cambridge
Department of Computer Science and Technology, William Gates Building, 15 J J Thomson Ave, Cambridge CB3 0FD, United Kingdom, ses88@cam.ac.uk

(2) University of Cambridge
MRC Cognition and Brain Sciences Unit, 15 Chaucer Road, Cambridge CB2 7EF, United Kingdom, Olaf.Hauk@mrc-cbu.cam.ac.uk

(3) University of Cambridge
MRC Cognition and Brain Sciences Unit, 15 Chaucer Road, Cambridge CB2 7EF, United Kingdom, Rezvan.Farahibozorg@mrc-cbu.cam.ac.uk

Keywords: EEG, MEG, language, support vector machine, semantics.

Abstract. How the brain processes language is not completely understood. Alternative hypotheses exist which attempt to explain the phenomenon. In order to provide a statistical analysis on the latency of deciphering semantic categories in the human brain, we used Electroencephalography (EEG) and Magnetoencephalography (MEG) data from 17 subjects, reading abstract or concrete words. We have applied support vector machine classification to differentiate between the brain states stemming from processing the two word categories. Our study suggests that the latency from semantic processing is in the ∼80-250ms range with EEG data giving the earliest indication. We found that brain state separability was most easily discernible from MEG sensors and reaches a maximum at ∼400-600ms. We hypothesize the early effects reflect semantic information retrieval, while the later effect may reflect mental imagery or decision-related processes.

1 Scientific Background

There is significant existing work in studying the neuronal mechanism underpinning language processing. Many studies conclude there is high spatial complexity in comprehending words [1, 2, 3]. Hence, it has become evident that phonological and semantic interpretation takes place in a distributed network of cortical regions with temporal dependencies between them [4, 5, 6]. Current agreement on the topic is that the left frontal and temporal cortices, namely Broca's and Wernicke's regions, play a principal role in word meaning comprehension. What remains controversial is the processing importance of other areas, as well as the temporal contingencies and communication in the whole neuronal network. This study aims to provide a more rigorous statistical analysis on the time course of semantic processing in the brain.

Many studies in neuroscience generally employ neuroimaging techniques like Functional Magnetic Resonance Imaging (fMRI), Positron Emission Tomography (PET), or scalp-recording techniques, such as Electroencephalography (EEG) and Magnetoencephalography (MEG). These approaches introduce both challenges and advantages. Modalities such as fMRI offer spatial resolution in the millimeter range but hinder researchers with their unsatisfactory performance in the time domain. EEG and MEG excel in the area of high millisecond resolution at the expense of a loss in spatial resolution. EEG and MEG measure the electric and magnetic fields on the scalp, respectively, stemming from brain activity. Both monitoring methods have tens of electrodes positioned at specific locations.

Two goals are achieved in this study. Firstly, we exploit the temporal advantage of EEG and MEG to improve the understanding of the time course of semantic word processing with a focus on abstract and concrete concepts. We use Multivariate Pattern Analysis (MVP) to discriminate between brain states stemming from processing different word categories. A similar approach was successfully utilized by Cichy et al. [7] to investigate whether visual information is encoded by cortical columns in the brain. We also show how we can enhance MVP classification accuracy by performing feature engineering so as to boost brain state separability. We then use the improved classification rate to identify the earliest point in time when the accuracy rise can be considered statistically significant. This latency is equivalent to the brain understanding word meaning. A major achievement is providing statistical assessment of how long semantic interpretation lasts. Secondly, we analyze which sensor types contain the most discriminative information in word meaning. This analysis is conducted both with respect to maximum classification accuracy and latency.

2 Materials and Methods

We use 17 subjects and observe their reactions to seeing abstract and concrete words in a random succession on a screen. The words are presented randomly so that the subjects do not know what word (and its category) will be shown next. The evoked potentials from brain activity in response to each word trial, i.e. an epoch, are then recorded (see fig. 1). To determine the time course of word processing in the brain MEG and EEG signal data was analyzed with a Support Vector Machine (SVM). The libsvm implementation by Chih-Chung Chang and Chih-Jen Lin was used [10]. The sensor data was collected in trials of 85 and 95 abstract and concrete words, respectively. An overall of 376 sensors were used during testing - 102 magnetometers, 204 gradiometers (overall 306 MEG sensors) and 70 EEG sensors.

All epochs were sampled at 1000Hz from 500ms before a subject had seen a word to 700ms after. Bad EEG channels (exceeding a voltage threshold) were removed and estimated using the average of nearby channels. A MaxFilter was used to remove far-away artefacts from the MEG signals. All epoch waveforms also underwent the following conditioning: firstly, they were FIR filtered between 1 Hz and 45 Hz; the signal baseline from -500ms to 0s was subtracted from each time series, and finally electrooculography (EOG) components and other noise (such as mains interference) were eliminated. The resulting time series were then down-sampled to 200Hz to speed up processing. Feature-scaling and normalization were also employed.

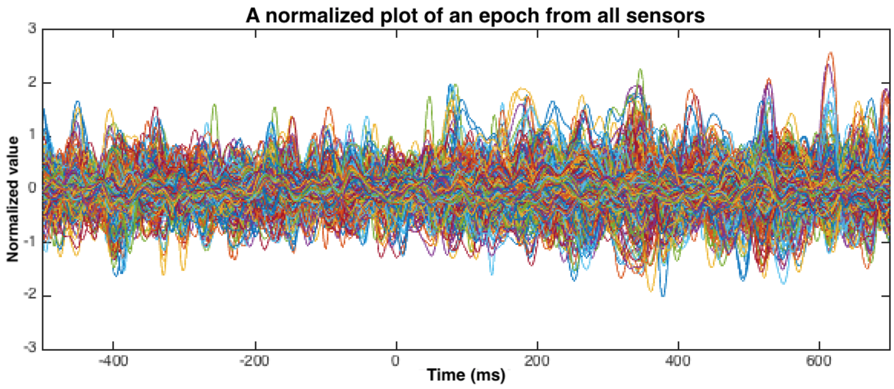

Figure 1: Visual representation of a word epoch from a single subject. The amplitude of the evoked potentials has been normalized. It can be seen that the amplitude response of some channels rises after 0ms, i.e. after the subject sees a word on the screen, but there is no obvious trend.

To improve the signal-to-noise ratio prior to classification we averaged several epochs from the same word category in order to create more ”typical” samples of the respective class. By utilizing this approach we can engineer features with more pronounced behaviour and make averaged epochs of concrete and abstract trials appear more different to the classification algorithm. To achieve this M random epochs were sampled without replacement and averaged from a word category. The procedure was repeated until all epochs were exhausted.

The resulting dataset was split in training and validation (75%), and test (25%) sets. The SVM classifier was trained with a linear kernel for each time index. We performed 5-fold cross-validation for hyperparameter optimization. The classifier was applied on the test set to create a timeline accuracy of word category discrimination for various sensor groups - magnetometers, gradiometers, EEG and combined. The whole procedure was repeated 10 times for each subject and the average test classification was reported for analysis.

T-testing was used to determine statistical significance in the time course analysis of semantic word processing. The distribution of accuracy timelines across all subjects was tested against a null hypothesis of 50% accuracy. Statistical significance was considered at the 5% level. As false positives can arise by chance, false discovery rate [9] and Bonferroni correction [8] were employed to account for any significant t-values stemming from random effects.

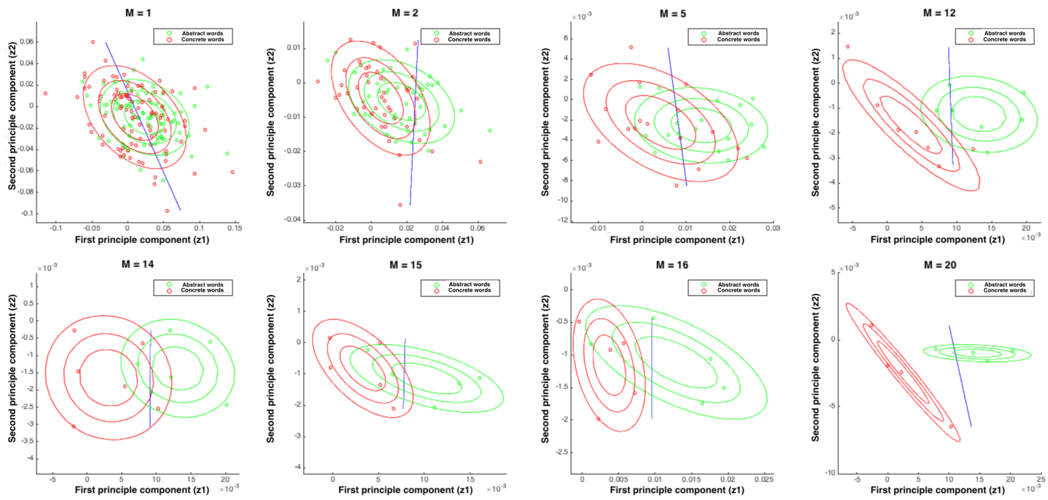

Figure 2: The figure represents 2D embeddings of abstract and concrete epochs of a single subject. M is the number of epochs averaged without replacement as feature engineering. It is evident that as M increases, the variances of the empirical word distributions decrease which improves class separation. The drawback of this feature engineering approach via averaging is that the number of training samples in the set decreases which hampers generalization error. This is best exemplified when $M = 20$ as the SVM separating hyperplane (depicted as a blue line in the figures) is not positioned well between the two classes.

3 Results

3.1 Determining the time course of semantic word processing using SVM classification

The aim of this experiment is to estimate the semantic processing latency period in human subjects. Fig. 3 depicts the timeline accuracy of word category discrimination we obtained by applying a SVM on data from magnetometers, gradiometers, EEG and combined sensor inputs. There are two key observations we can make. Firstly, it is expected that the classification accuracy prior to 0ms should be 50% as the subjects have not been shown a word yet. Secondly, there is significant noise in the time course accuracy signals. Despite this, it can be seen that an accuracy peak is achieved between $\sim$ 400ms and 600ms.

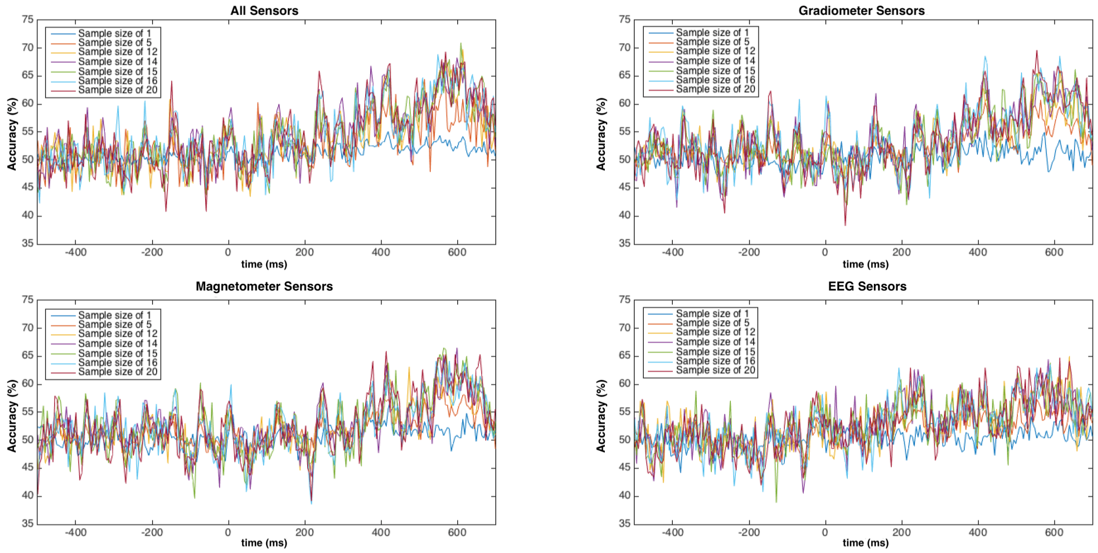

Figure 3: The average time course accuracy between word processing brain states across all subjects for various feature engineering sample sizes and sensor types. A pronounced accuracy peak is achieved between $\sim$ 400ms and 600ms. The signals are too noisy to visually identify when word meaning is initially deciphered. The sample sizes refer to the number of averaged epochs in the feature engineering procedure.

Employing a t-test with a null hypothesis of 50% accuracy at the 5% significance level gave many rejected null hypotheses prior to 0ms. Since the subjects were shown words from the two categories randomly, and could not have expected what words or its category is coming next, we concluded the rejected null hypotheses prior to 0ms to be false positives. Consequently, we applied the false discovery rate (FDR) method to limit the expected proportion of discoveries (rejected null hypotheses) that are false (also at the 5% level). This procedure gave rise to a new set of FDR-adjusted p-values shown in fig. 4. The results reveal pronounced clusters of p-values at $\sim$ 400ms and $\sim$ 600ms corresponding to the increase in brain state separability we observed in fig. 3 in the same time frame. We hypothesize that this phenomenon is indicative of later mental imagery or decision-related processes.

Identifying the earliest signs of brain state separability proved to be a very difficult task. Even the FDR-controlled p-values from all sensor type inputs but the EEG data show false discoveries prior to 0ms. To obtain an upper bound on semantic processing latency we applied the conservative Bonferroni correction which limits the probability of having even one false discovery. We set the level of control such that there are no Bonferroni-adjusted p-values prior to 0ms. In this case the earliest rejected null hypothesis occurs at $\sim$ 290ms. This conservative estimate is corroborated by a group of FDR-adjusted p-values from all sensors at $\sim$ 235ms and also FDR-adjusted p-values from EEG sensor data at $\sim$ 200ms-250ms (see fig. 4). On the other hand, a more optimistic estimate of semantic word processing latency can be obtained by taking the first FDR-adjusted p-value after 0ms, which occurs at $\sim$ 80ms for results from EEG and all sensors data. Consequently, our data analysis suggests that semantic processing in the brain occurs approximately $\sim$80ms-250ms after a word is shown on the screen.

3.2 Comparing classification accuracy for different sensor types

Visual inspection of fig. 3 reveals that MEG contributes more to maximum accuracy classification than EEG. This enabled us to observe the later accuracy peaks between $\sim$ 400ms and 600 ms. On the other hand, EEG data gave us the earliest indication of brain state separability. In addition, FDR-adjusted p-values based on EEG did not suffer from false discoveries prior to 0ms.

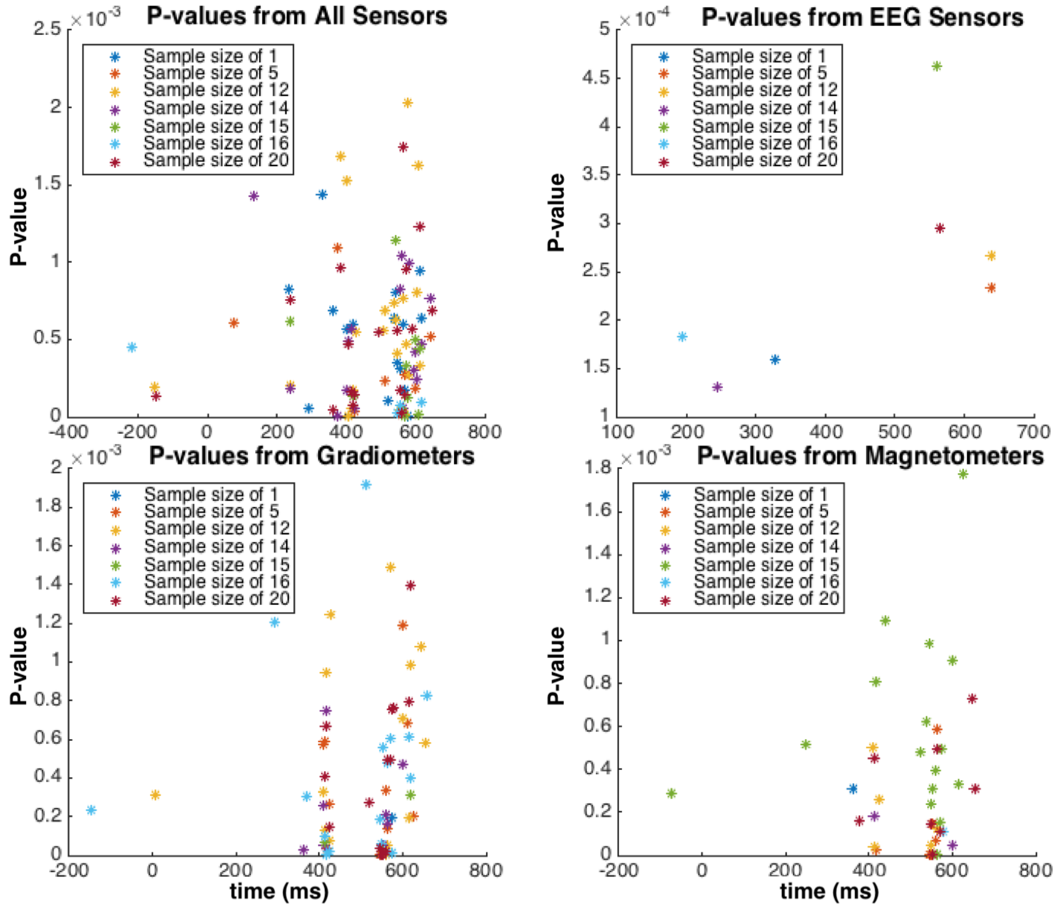

Figure 4: A plot of FDR$\leq$ 5% adjusted p-values showing rejected null hypotheses. The clusters between $\sim$ 400ms and 600ms correspond to the increased accuracy in the same time period in fig. 3. The sample sizes refer to the number of averaged epochs in the feature engineering procedure.

4 Conclusion

We were able to use support vector machine classification, and EEG and MEG data to discriminate between brain states, corresponding to the semantic processing of abstract and concrete word categories. We used statistical testing, and false discovery rate and Bonferroni correction procedures to provide a statistical view on the latency of word meaning decoding in human subjects. Our results suggest that EEG sensors give the earliest indication of semantic processing which occurs $\sim$80ms-250ms after a word is shown to the subject. MEG sensors provided the best data for brain state separability with maximum classification accuracy $\sim$400ms-600ms after a word is shown to the subject.

Acknowledgments

I would like to thank Dr. Adrian Weller, senior research fellow in machine learning at the University of Cambridge, for his expert advice and guidance.

This research was supported by the Engineering and Physical Sciences Research Council [EP/L015889/1].

References

[1] Dronkers, N.F., Wilkins, D.P., Valin, R.D.V., Redfern, B.B., Jaeger, J.J., 2004. *Lesion analysis of the brain areas involved in language comprehension.*, Cognition 92, 145177. doi:10.1016/j.cognition.2003.11.002

[2] Dronkers, N.F., Plaisant, O., Iba-Zizen, M.T., Cabanis, E.A., 2007. *Paul Broca's historic cases: high resolution MR imaging of the brains of Leborgne and Lelong.* Brain 130, 14321441. doi:10.1093/brain/awm042

[3] Belin, P., Zatorre, R.J., Ahad, P., 2002. *Human temporal-lobe response to vocal sounds.* Cognitive Brain Research 13, 1726. doi:10.1016/s0926-6410(01)00084-2

[4] Friederici, A.D., Pfeifer, E., Hahne, A., 1993.*Event-related brain potentials during natural speech processing: effects of semantic, morphological and syntactic violations.* Cognitive Brain Research 1, 183192. doi:10.1016/0926-6410(93)90026-2

[5] Hauk, O., Johnsrude, I., Pulvermller, F., 2004. *Somatotopic representation of action words in human motor and premotor cortex.* Neuron 41, 301307. doi:10.1016/s0896-6273(03)00838-9

[6] Hauk, O., Pulvermller, F., 2004.*Neurophysiological distinction of action words in the fronto-central cortex.* Human Brain Mapping 21, 191201. doi:10.1002/hbm.10157

[7] Cichy, R.M., Ramirez, F.M., Pantazis, D., 2015. *Can visual information encoded in cortical columns be decoded from magnetoencephalography data in humans?*, NeuroImage 121, 193204. doi:10.1016/j.neuroimage.2015.07.011

[8] Bland, J.M., Altman, D.G., 1995.*Statistics notes: Multiple significance tests: the Bonferroni method.* BMJ 310, 170170. doi:10.1136/bmj.310.6973.170

[9] Y. Benjamini, Y. Hochberg *Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing*, Journal of the Royal Statistical Society, Series B (Methodological), Vol. 57, No.1 (1995), pp. 289-300

[10] Chih-Chung Chang and Chih-Jen Lin, *LIBSVM : a library for support vector machines.* ACM Transactions on Intelligent Systems and Technology, 2:27:1–27:27, 2011. Software available at http://www.csie.ntu.edu.tw/ cjlin/libsvm