April 20, 2017
This is a preprint of an article submitted to Journal of Cognitive Neuroscience

Exploring the organization of semantic memory through unsupervised analysis of event-related potentials

Marijn van Vliet[1,2*], Marc M. Van Hulle[2], and Riitta Salmelin[1]

[1]Aalto University, Department of Neuroscience and Biomedical Engineering, Espoo, Finland
[2]KU Leuven – University of Leuven, Department of Neurosciences, Laboratory for Neuro- & Psychophysiology, Leuven, Belgium
[*]Corresponding author: w.m.vanvliet@gmail.com

Abstract

Modern multivariate methods have enabled the application of unsupervised techniques to analyze neurophysiological data without strict adherence to pre-defined experimental conditions. We demonstrate a multivariate method that leverages priming effects to shed light on the organization of memory representations in the brain. The current study focuses on the semantic relationships that play a key role in the organization of our mental lexicon of words and concepts. The N400 component of the event-related potential is considered a reliable neurophysiological response that is indicative of whether accessing one concept facilitates subsequent access to another (i.e., one "primes" the other). To further our understanding of the organization of the human mental lexicon, we propose to utilize the N400 component to drive a clustering algorithm that can uncover, given a set of words, which particular sub-sets of words show mutual priming. Such a scheme requires a reliable measurement of the amplitude of the N400 component without averaging across many trials, which was here achieved using a recently developed multivariate analysis method based on beamforming. We validated our method by demonstrating that it can reliably detect, without any prior information about the nature of the stimuli, a well-known feature of the organization of our semantic memory: the distinction between animate and inanimate concepts. These results motivate further application of our method to data-driven exploration of disputed or unknown relationships between stimuli.

Keywords: event-related potential, EEG, N400, semantic clustering, reading

1 Introduction

Semantic priming experiments[1] have revealed that accessing a word in our mental lexicon facilitates future access to semantically related words. Since words usually occur in a logical sequence, this "priming" behavior facilitates the processing of likely continuations of a sentence or story[2] and thereby contributes to our ability to exchange messages with others at high speed.

[1] McNamara and Holbrook, 2003; Neely, 1991

[2] Neely, 1976

The semantic priming effect has been helpful for studying the organization of human semantic memory.[3] For example, the exact nature of the relationships that causes one word to prime another word continues to be the focus of research.[4] In this paper, we demonstrate how unsupervised techniques, such as hierarchical clustering, are a particularly useful tool in this case and develop a new technique to study the organization of semantic memory based on a neural correlate of the semantic priming effect.

[3] e.g. Collins and Loftus, 1975; Kutas and Federmeier, 2000

[4] e.g. De Deyne, Navarro, and Perfors, 2016; C. K. Van Petten, 1993

The boost in signal-to-noise ratio (SNR) provided by multivariate data analysis[5] enables an exciting paradigm shift in how new insights may be obtained from neu-

[5] Friston et al., 1996; Norman, Polyn, Detre, and Haxby, 2006

rophysiological data. When the SNR is high enough, a researcher can approach the data analysis in an unsupervised manner, instead of labeling data according to some predetermined division (e.g., words vs. pseudowords or tools vs. vegetables). Multivariate analysis reduces the need for averaging across trials, thus facilitating the generation of sufficiently many data points for learning the underlying structure in the data distribution, for example via clustering techniques.[6] This allows for a data-driven approach to complement theoretical work.

[6] Jain, Murty, and Flynn, 1999

In the application of clustering techniques, the key component to consider is the (dis)similarity score employed by the algorithm. This score is a measure of the distance between two items and is used by the clustering algorithm to determine which items to group together in a cluster. Hence, the effectiveness and validity of clustering techniques in neuroscience depends a great deal on how the measured brain activity is translated into a similarity score.

In the context of semantic relationships, the similarity score corresponds to the concept of semantic distance.[7] Such distance metrics are traditionally based on behavioral data, such as the co-occurrence of words in a large text corpus,[8] degree of overlap of semantic features[9] or the forward association strength (FAS) score which is produced by performing an association study where participants, presented with a target word, are asked to write down which words come to mind.[10] These metrics are all based on data that was produced by participants who were given enough time to consciously think about their responses. However, for the purposes of this study, a semantic distance metric is preferred that is based on a neurophysiological response that occurs while the target word is being processed, before any conscious decision making process can be completed.

[7] Rips, Shoben, and Smith, 1973

[8] Jones, Willits, and Dennis, 2015

[9] De Deyne and Storms, 2008; Hutchison, 2003; McRae, Cree, Seidenberg, and McNorgan, 2005

[10] De Deyne, Navarro, and Storms, 2013; Nelson, McEvoy, and Schreiber, 2004

Previous studies that have developed semantic distance metrics from brain activity did so by showing that concepts belonging to the same natural semantic category (e.g., tools, animals, etc.) produce similar brain activity. For example, functional magnetic resonance imaging (fMRI) studies have shown that stimuli from the same semantic category generate similar blood-oxygen-level dependent (BOLD) activity patterns[11] and electroencephalography (EEG) and magnetoencephalography (MEG) studies have shown that they produce similar spatio-temporal time courses.[12] However, while some semantic categories may activate unique brain activity patterns, there is currently no consensus that this should be the case for all categories[13] or, for that matter, other types of relationships that are important to the semantic systems in our brain. In this study, we explore an alternative route to obtain a semantic distance metric that is more closely tied to semantic priming.

[11] Gerlach, 2007; Huth, De Heer, Griffiths, Theunissen, and Jack, 2016, 2012

[12] Chan, Halgren, Marinkovic, and Cash, 2011; Simanova, van Gerven, Oostenveld, and Hagoort, 2010

[13] Pulvermüller, 2013

The distance metric employed in this study is based on an component of the event-related potential (ERP) as recorded through EEG, which has been shown to be reliably modulated by semantic priming. By contrasting different levels of priming, an effect can be seen that reaches its maximum around 400 ms post stimulus onset and the component was hence named the N400.[14] Since its discovery, relative changes in the amplitude of the N400 component have been shown to correlate well with various behavioral metrics of the strength of the semantic relationship between words, such as word co-occurrence,[15] FAS[16] and semantic feature overlap.[17]

[14] Kutas and Federmeier, 2011; Kutas and Hillyard, 1984

[15] C. Van Petten, 2014

[16] Luka and Van Petten, 2014; van Vliet et al., 2016

[17] Koivisto and Revonsuo, 2001

In this study, we demonstrate how to find semantic clusters for a given set of words

by measuring the amplitude of the N400 component that was evoked in a semantic priming experiment. Since the semantic priming effect and its relation to the N400 component have been thoroughly studied, the metric and the clustering result it produces are straightforward to interpret.

EEG was recorded while all pairwise combinations of the stimuli, a set of 14 written words, were presented sequentially to the participants. For the second word of each word-pair (the target), the amplitude of the N400 component of the evoked EEG response was estimated using an linearly constrained minimum variance (LCMV) beamformer,[18] modified to be suitable for ERP analysis.[19] The resulting N400 amplitudes formed the elements of a word-to-word distance matrix that served as input to a hierarchical clustering algorithm, with the aim to discover clusters of semantically related words. Since the main focus of this study is to explore if such a scheme can work, the chosen stimuli in this study were either animals or furniture items, thus items that most semantic theories place in separate clusters.[20] The validity of the method was assessed by determining whether the clustering algorithm reveals these clusters.

[18] Van Veen, Van Drongelen, Yuchtman, and Suzuki, 1997

[19] Treder, Porbadnigk, Shahbazi Avarvand, Müller, and Blankertz, 2016; van Vliet et al., 2016; Wittevrongel and Van Hulle, 2016

[20] Martin, 2007

Importantly, even though the stimuli in this study were designed with a clear dichotomy, the method will be agnostic to this fact. Accordingly, the proposed method should also be suitable for exploring datasets where the proper clustering is ambiguous or disputed. Furthermore, due to the unsupervised nature of the method, additional sub-clusters may also be revealed that were not an intentional part of the experimental design.

2 Methods

The study was performed with 19 participants. The data of two participants was discarded due to poor sensor contact quality and the data of one participant was discarded due to excessive eye blinks. Of the remaining 16 participants, 10 were male and 6 female, in the age range of 20 to 58 years (mean 38, std 11 years), all but one were right handed, 6 were native speakers of Walloon-French and the other 10 native speakers of Flemish-Dutch.

This study was performed at KU Leuven and ethical approval was obtained from its university hospital's medical ethics committee. All participants were unpaid volunteers who signed an informed consent form before the experiment.

2.1 Stimuli and experimental procedure

Word-pairs were formed by using all possible prime–target combinations (182) of the 14 words listed in table 1. The list contains category exemplars for African animals and common furniture items. The stimuli differ in length and frequency of usage, which are normally controlled for in linguistic experiments. However, our method is mostly insensitive to the influences of such word-specific properties, as will be further argued in the Discussion section. The stimuli were presented in the native language of the participant (Flemish-Dutch or Walloon-French). All possible word-pairs were presented once, which means that each individual word was presented

Dutch	French	English
bed	lit	bed
bureau	bureau	desk
deur	porte	door
giraf	girafe	giraffe
kast	placard	closet
leeuw	lion	lion
neushoorn	rhinoceros	rhinoceros
nijlpaard	hippopotame	hippopotamus
olifant	éléphant	elephant
stoel	chaise	chair
tafel	table	table
tijger	tigre	tiger
zebra	zèbre	zebra
zetel	canapé	couch

Table 1: Words used in the unsupervised clustering study. The words were displayed in French or in Dutch, according to each participant's native language. The English translation is only for the sake of exposition and was not displayed to the participants. The stimuli consisted of all possible pairwise combinations of these words.

26 times: 13 times as prime and 13 times as target.

Participants were seated in an upright position approximately one meter from a computer screen. The hand used to give the button response rested upon a table with the index and middle fingers on the mouse buttons. A trial consisted of the sequential presentation of a single word-pair. The first word of the word-pair (the prime) was presented for 200 ms and the second word (the target) for 1000 ms with a stimulus onset asynchrony of 500 ms, after which a question mark appeared prompting a response.

Following the advise of Renoult and Debruille (2011) for obtaining a semantic priming effect even when stimuli are shown multiple times during the experiment, the participants were asked to determine whether the cue and target words belonged to the same semantic category by pressing one of two mouse buttons. The mapping of the yes/no response to the mouse buttons and the hand used to operate the mouse were counterbalanced independently across participants.

2.2 Data recording and preprocessing

EEG was recorded continuously using 32 active electrodes (extended 10–20 system) with a BioSemi Active II System (BioSemi, Amsterdam, the Netherlands), having a 5th order frequency filter with a pass band of 0.16 Hz to 100 Hz, and sampled at 2048 Hz. Two additional electrodes were placed on both mastoids and their average signal was used as a reference for the other sensors. Furthermore, four additional electrodes were placed on the outer canthi of the eyes and above and below the left eye to record a horizontal and vertical electro-oculogram (EOG).

The EEG and electrooculogram (EOG) signals were further bandpass filtered offline between 0.3 Hz to 30 Hz by a 4^{th} order zero-phase infinite impulse response (IIR) filter to attenuate large drifts and irrelevant high frequency noise. Electrodes with insufficient signal quality were detected based on visual inspection of the raw data and replaced by a virtual channel using spherical interpolation of the remaining electrodes.[21] The EOG signal was used to attenuate eye artifacts from the EEG signal using the aligned-artifact average regression method described in.[22] Individual

[21] Perrin, Pernier, Bertrand, and Echallier, 1989

[22] Croft and Barry, 2000

trials were obtained by cutting the continuous signal from 0.1 s before the onset of each target stimulus to 1.0 s after. All trials were used in the analysis. Baseline correction was performed using the average voltage in the 0.1 s interval before the stimulus onset as baseline value. Finally, since any high frequency content was removed by the band pass filter, the signal was downsampled to 50 Hz without losing much information. This step was included to reduce the dimensionality of the data matrices, which improves the numerical stability of the beamformer filter.

2.3 Beamformer filter

After preprocessing the EEG signals, multivariate analysis was performed using a spatio-temporal LCMV beamformer filter. The filter takes a weighted sum of the data points from all EEG channels and all samples within an epoch. The result of this summation represents the estimated amplitude of the N400 component of the ERP within that epoch. For an in depth explanation and implementation details of the method, see van Vliet et al. (2016).

The beamformer approach consists of two steps. The first step is to construct a template of the desired signal: in this case the spatial and temporal shape of the N400, which is derived with traditional ERP analysis and consists of averaging many epochs across many subjects. The second step is to obtain the set of weights that isolates this signal from the rest of the EEG, which entails estimating the inverse signal covariance matrix of the recording currently under consideration (the target recording). The advantage of this approach is that data recorded during previous studies can be re-used to acquire the N400 template, which removes the requirement for the target recording to have predefined experimental conditions, i.e., indicating beforehand which trials are assumed to have a high or low N400 amplitude.

To obtain a template of the N400 component and fine-tune the beamformer filter, we re-used data that was collected in a previous semantic priming study.[23] In this study, 10 native speakers of Flemish-Dutch were shown 800 word-pairs with varying FAS, as determined from an association norm database compiled by De Deyne and Storms,[24] covering the whole range of completely unrelated to the strongest related words in the database. The experimental procedure, recording setup and data processing were identical to the one used for the unsupervised clustering study as described above, with the exception that the responding hand was always the right hand and the mapping of yes/no responses to the mouse buttons was not counterbalanced. See van Vliet et al. (2014) for further details about the study.

[23] van Vliet et al., 2014

[24] De Deyne et al., 2013; De Deyne and Storms, 2008

The data of the previous study was re-analyzed by performing linear regression, using the logarithm of the FAS of the stimuli as predictor and the EEG as response variable, resulting in what Smith and Kutas (2015) refer to as a "slope" ERP. This slope ERP is a generalization of the difference wave and can be thought of as "the part of the ERP that changes when the FAS of the stimulus changes." Next, we determined the time point when the global field power (GFP) of the slope ERP reached its maximum, which was at 430 ms after stimulus onset. The distribution of the slope ERP across the sensors at that time point was taken to be the spatial pattern for the N400 component (figure 1, left).

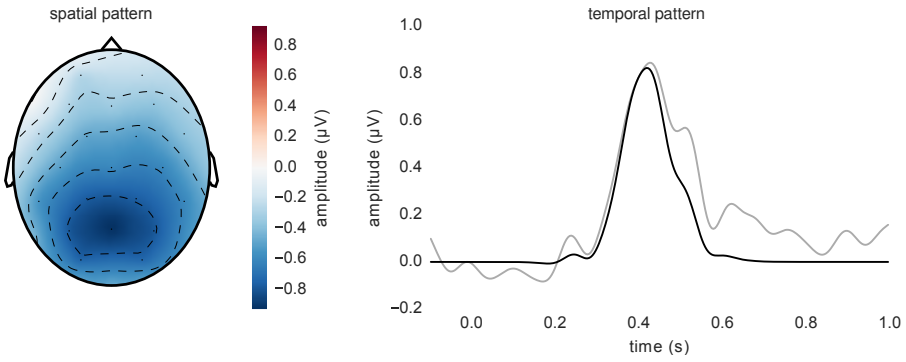

Figure 1: Spatial pattern (left) and temporal pattern (right) of the N400 component of the ERP, evoked in a semantic priming experiment. In the figure depicting the temporal template, the gray line represents the result of the spatial beamformer and the black line represents the result after multiplying with a Gaussian kernel.

The temporal template was constructed by using the spatial template to create a spatial LCMV beamformer,[25] the output of which represents an estimation of the summed activity at the cortical source locations of the N400 (figure 1, right, gray line). This time course was further refined by multiplying it with a Gaussian kernel ($\mu = 400\,\text{ms}$, $\sigma = 0.1\,\text{ms}$) which has the effect of limiting the non-zero values to a window of interest centered around the peak amplitude of the N400 (figure 1, right, black line). Finally, the full spatio-temporal template was obtained by taking the outer product of the spatial and temporal templates.

[25] van Vliet et al., 2016

To compute the filter weights that will isolate the signal component described by the template from the rest of the signal, the template must be multiplied with the spatio-temporal covariance matrix Σ of the target data. This matrix can be readily computed from the data of the current study, since it does not require contrasting different experimental conditions. Due to the high dimensionality of this matrix, it is recommended to employ heavy shrinkage during its estimation. In this study, we employed shrinkage towards the diagonal:

$$\hat{\Sigma} = \mathbf{X}\mathbf{X}^{\mathsf{T}}, \tag{1}$$

$$\Sigma = (1-\alpha)\hat{\Sigma} + \alpha\frac{\mathrm{Tr}\hat{\Sigma}}{n}\mathbf{I}, \tag{2}$$

where $\mathbf{X}$ is a matrix where each row corresponds to one of the n epochs and contains a flattened version (i.e. all elements are placed on a single row) of the (channels × samples) matrix. $\hat{\Sigma}$ is the empirical covariance matrix, "Tr$\hat{\Sigma}$" means the sum of the diagonal elements of $\hat{\Sigma}$, and $\mathbf{I}$ is an identity matrix.

The value of the shrinkage parameter α was optimized by designing beamformer filters with different values for α and applying them to the data of the previous study. The optimization criterion was to maximize the correlation between the output of the filter and the FAS of the stimuli that were used in that study. This resulted in an optimal α value of 0.9, which is the value we subsequently used to design the filter for the present study.

Given the covariance matrix and the template of the N400 component, the estima-

tion of the amplitude of this component ($\hat{y}$) for a given epoch is:

$$\mathbf{w} = \frac{\Sigma^{-1}\mathbf{a}}{\mathbf{a}^{\mathsf{T}}\Sigma^{-1}\mathbf{a}}, \tag{3}$$

$$\hat{y} = \mathbf{w}^{\mathsf{T}}\mathbf{x}, \tag{4}$$

where Σ^{-1} is the inverse of the covariance matrix, $\mathbf{a}$ is a flattened version of the (channels × samples) matrix containing the N400 template and $\mathbf{x}$ is the flattened version of the (channels × samples) matrix containing the EEG epoch.

2.4 Hierarchical clustering

The amplitude of the N400 ERP component $\hat{y}$, as quantified by the spatio-temporal LCMV beamformer filter, was further processed to obtain a suitable metric for the semantic distance between the prime and target stimuli. For each participant, z-scoring was performed across the $\hat{y}$'s in order to equalize the scalings. The z-scored N400 amplitude estimates were subsequently organized in a (words × words) matrix $\mathbf{D}$.

Since we are interested in the N400 *effect*, i.e. the relative amplitude difference of the component across contrasting conditions, each column of $\mathbf{D}$ was normalized by removing its mean. This has the effect of removing the baseline N400 response to each word and only preserving relative changes in N400 amplitude as the target word is presented in combination with different cue words.

Matrix $\mathbf{D}$ contains, for each pairwise combination of two words in table 1, two responses for each participant. One response for the case where the first word was used as cue and the second word as association and another response for the reversed case. Since the hierarchical clustering algorithm operates on geometric distance, which is symmetric and positive, the distance matrix $\mathbf{D}$ should be symmetric and positive as well. This was achieved by averaging $\mathbf{D}$ with its transposed form and subtracting the lowest value:

$$\mathbf{D}_{\text{sym}} = \frac{\mathbf{D} + \mathbf{D}^{\mathsf{T}}}{2}, \tag{5}$$

$$\mathbf{D}_{\text{pos}} = \mathbf{D}_{\text{sym}} - \min \mathbf{D}_{\text{sym}}. \tag{6}$$

The final matrix used as input for the hierarchical clustering algorithm was obtained by averaging the distance matrices across participants. Since the N400 amplitude estimates are noisy, it is beneficial to base the distance between two clusters on as many measurements as possible. Therefore, average linkage (also known as unweighted pair group methods with arithmetic mean (UPGMA) linkage) was chosen as the clustering algorithm.[26] It determines the distance between two clusters by considering the average distance between all items in the clusters.[27] We present the output of the clustering algorithm in the form of a dendrogram.

[26] Jain et al., 1999

[27] Sokal and Michener, 1958

2.5 Statistics

At each "node" in the dendrogram, where two sub-clusters are joined together to form a new cluster, a statistical test was performed to provide an indication of the reliability of the distinction presented by the two sub-clusters. To this end, a linear mixed effects (LME) model was used to analyze the difference between the estimated N400 amplitudes in response to the within-cluster word-pairs versus the between-cluster word-pairs. Note that this test can only be performed if both clusters consist of at least two words, otherwise there are no within-cluster word-pairs. The normalized N400 amplitudes (the elements of the asymmetric matrix **D**) were used as the dependent variable, with a dummy encoding of the labels "within-cluster" (= 1) versus "between-cluster" (= 0) as fixed effect. Since the model needs to generalize beyond the participants included in the study, participants were modeled as random effect (random slopes and random intercepts). However, since the model does not need to generalize beyond the words in the clusters, words were not included as random effect. The model was fitted using restricted maximum likelihood (REML), with degrees of freedom and the resulting p-values estimated using Satterthwaite's approximation.[28] To control for family-wise error rate (FWER), the p-values were Bonferroni corrected my multiplying them by the number of tests performed. When this resulted in $p > 1$, we report $p = 1$.

[28] Satterthwaite, 1946

2.6 Software

Stimulus presentation was performed using MATLAB in combination with the Psychophysics toolbox.[29] Data analysis was performed using Python in combination with the Psychic, NumPy and SciPy packages.[30] Covariance estimation with shrinkage was performed using the Scikit-learn package.[31] Plots were created using the Matplotlib package.[32] Statistical analysis was performed using R[33] in combination with the LME4[34] and lmerTest[35] packages.

A software implementation of the N400 template estimation procedures and spatio-temporal LCMV beamformer can be found at:
https://github.com/wmvanvliet/ERP-beamformer.

[29] Brainard, 1997

[30] Oliphant, 2007

[31] Pedregosa et al., 2012

[32] Hunter, 2007

[33] R Core Development Team, 2015

[34] Bates, Maechler, Bolker, and Walker, 2015

[35] Kuznetsova, Brockhoff, and Christensen, 2015

3 Results

As expected, the button responses collected during the experiment showed that the participants very consistently marked word-pairs as "related" and "unrelated" according to a classification of animal versus furniture item. Furniture–furniture pairs received a "related" response in 89.0 % of the time, animal–animals pairs 93.6 %, furniture–animal pairs 1.1 % and animal–furniture pairs 0.6 %. It is likely that after a few trials, the participants noticed the pattern and started to perform a classification instead of a judgement of association task.

The distance matrix that was based on estimations of the amplitude of the N400 component (figure 2) also shows as overall trend a dichotomy between animal versus furniture items. Although single-item measurements can be unreliable (e.g. *chair–hippo* shows up as relatively related, which is probably a measurement error),

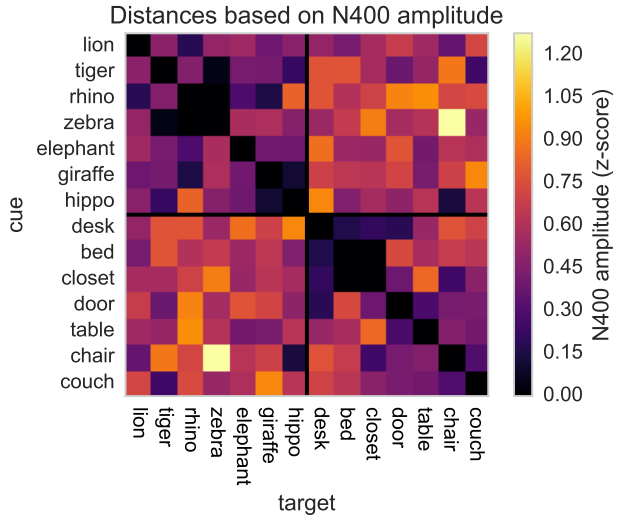

Figure 2: Distance matrix based on the amplitude of the N400 component, averaged across participants. The order of the words mirrors the order in which they appear in the dendrogram (figure 3). Black lines mark the boundary between the top clusters in the dendrogram.

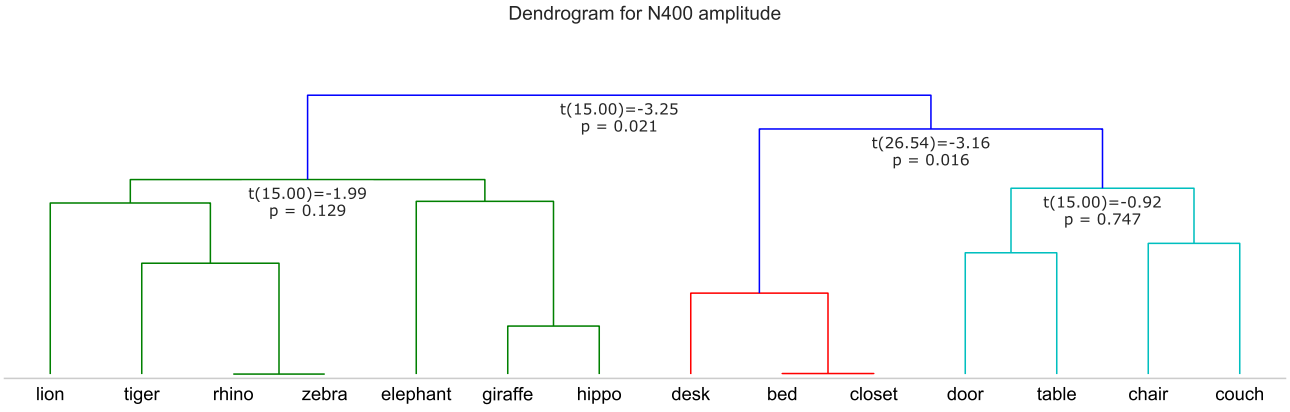

Figure 3: Dendrogram resulting from the hierarchical clustering algorithm applied to the distance matrix based on the amplitude of the N400 component. Statistical tests were performed to test for differences in N400 amplitude in response to between-(sub)cluster versus within-(sub)cluster word-pairs. Reported p-values are Bonferroni corrected.

hierarchical clustering can reveal the underlying patterns.

The dendrogram produced by the hierarchical clustering algorithm (figure 3) has as the topmost two clusters all the animal stimuli versus all the furniture stimuli. The fact that these clusters could be reliably reconstructed shows that the multivariate analysis of the EEG data yielded a measurement with a high enough SNR to perform this type of unsupervised clustering. As these clusters are themselves divided into sub-clusters, the results are based on less data and therefore less reliable. Statistical tests at each "node" of the dendrogram are an indication of this reliability and show whether there is a significant difference in N400 amplitude between within-cluster and between-cluster trials.

The only explicit distinction in the experimental design was a distinction between animals and furniture items. However, the dendrogram suggests that there may be a dichotomy in the chosen furniture stimuli. The cluster containing the animal stimuli did not show any reliable further sub-clustering.

The grand average ERPs, obtained by assigning the labels "within-cluster" and "between-cluster" based on the topmost clustering in the dendrogram, are presented in figure 4. Two components can be observed in the ERP, the first being the N400

component with a posterior distribution, present during both the within-cluster and between-cluster conditions. The second component is only observed in the between-cluster condition and has a more frontal distribution which can be possibly classified as a P600 component, commonly observed when stimuli are repeated.[36] Note that in the distance matrix (figure 2), the estimated N400 amplitudes were corrected by removing the mean along the columns in order to remove the "baseline" N400 response to each word. The ERP shown in figure 4 are uncorrected and therefore any visible effects are partly due to differences in the properties of the target words such as length, frequency, etc.

[36] Van Strien, Hagenbeek, Stam, Rombouts, and Barkhof, 2005

4 Discussion

The main result is that the distinction between animals and furniture items could be reliably extracted, based purely on EEG responses. This could be done without supplying any information about the nature of the clusters to the algorithm (i.e., no experimental conditions, no information about the clusters having an equal number of members), thus giving confidence that the method can produce trustworthy results for datasets where the optimal clustering is not known beforehand, provided that the distance (in our case semantic distance) between the clusters is large enough.

We employed a semantic distance metric that is based on the amplitude of the N400 component of the ERP, evoked using a semantic priming paradigm. This metric may capture different semantic relationships than earlier work that analyzed the full spatio-temporal activity pattern evoked by single words.[37] Furthermore, since the proposed metric does not require to distinguish brain activity between different spatial locations, the measurement can also be performed using techniques that have a relatively poor spatial resolution, such as EEG.

[37] Chan et al., 2011; Gerlach, 2007; Huth et al., 2016; Simanova et al., 2010

The method requires the detection of differences in N400 amplitude when a cue word is presented in combination with different prime words. How large these differences need to be in order for clusters to be differentiated depends on the SNR that can be achieved in estimating the amplitudes. In this study, we employed a spatio-temporal LCMV beamformer which has been shown to produce more reliable estimates of the N400 amplitude than more traditional approaches, such as measuring the mean voltage in a fixed time window.[38] The experimental paradigm used in this study adds some additional challenge, since stimuli need to be repeated in order to construct a full word-to-word distance matrix. Stimulus repetition is known to degrade the N400 effect due to semantic facilitation through short term memory (e.g., due to the old/new effect).[39] Nevertheless, our results reproduce the earlier finding that the N400 effect persists even when the stimuli are repeated,[40] as long as the target word cannot be predicted from the prime word and an explicit task is given to the participant.[41]

[38] van Vliet et al., 2016

[39] Rugg and Curran, 2007

[40] Debruille and Renoult, 2009; Renoult and Debruille, 2011

[41] Renoult, Wang, Calcagno, Prévost, and Debruille, 2012

The ability of the beamformer algorithm to accurately estimate N400 amplitudes depends greatly on the accuracy of the supplied template.[42] The fact that good results were obtained using a template based on an independent dataset (figure 1), provides some validation that the component reaching a maximum around 400 ms (figure 4)

[42] Treder et al., 2016; van Vliet et al., 2016

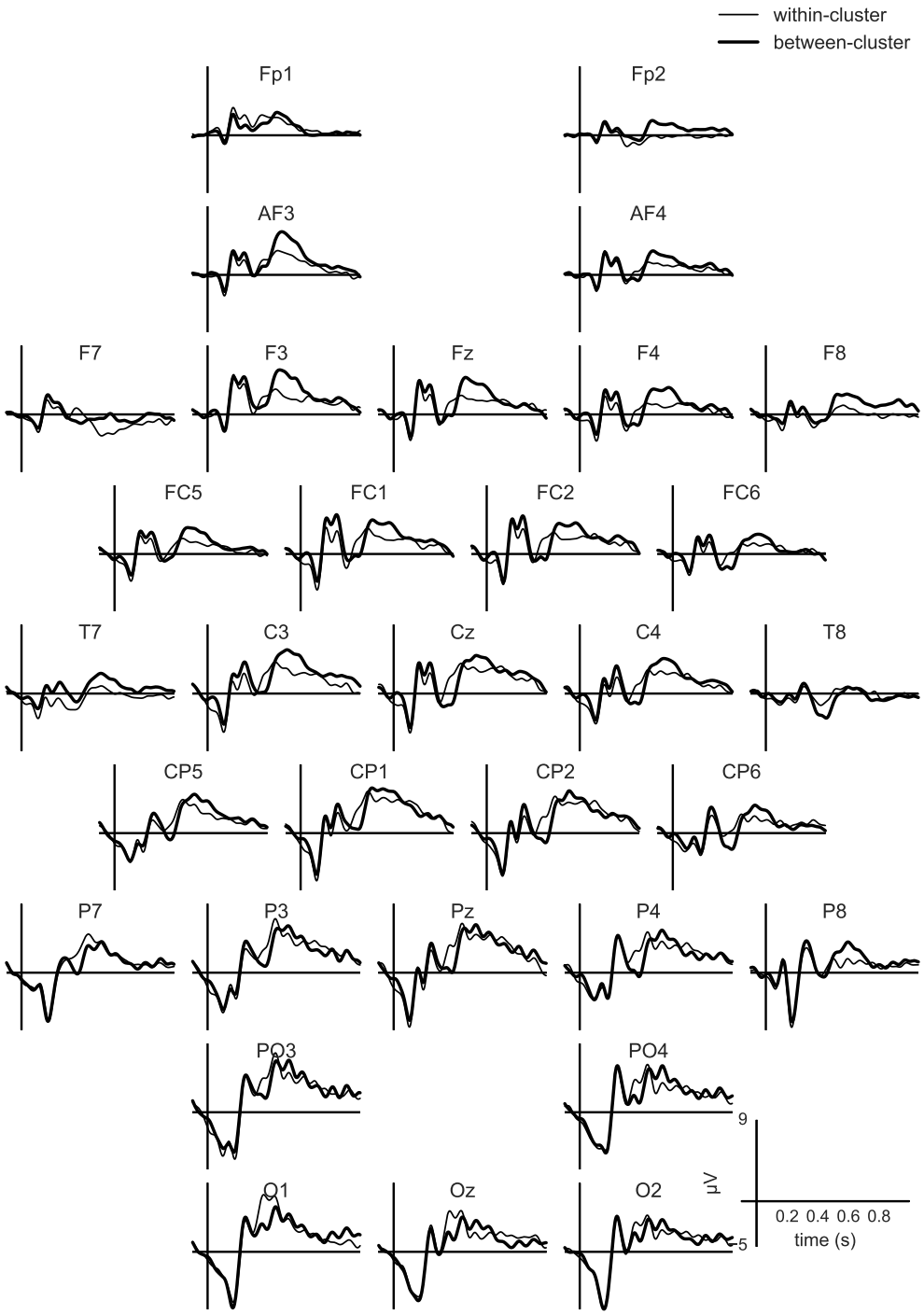

Figure 4: Grand-average ERPs in response to within-cluster (thin line) and between-cluster (thick line) word-pairs, corresponding to the top-most clusters in the dendrogram: animals versus furniture items.

is similar to the N400 component observed in classical priming experiments. If the component evoked in this study would deviate too much from the template (either in spatial distribution or timing), it would fall outside the passband of the filter. However, it is likely that there are small differences between the template and the N400 observed in this study, due to the repetition of stimuli, which can cause shifts in the timing of the component.[43] A template that is based on the data collected during this study is available upon request from the corresponding author, for use in future studies that employ a methodology that is similar to that in the current work.

[43] Renoult et al., 2012

It is worth noting that, although p-values are provided in the dendrogram, the clustering result goes beyond the statistical statement these p-values make. While there are many possible ways to cluster the stimuli in such a manner that there is a significant difference in N400 amplitude between the within-cluster and between-cluster pairs, the dendrogram reveals, out of all possible manners to arrange the items, the strongest hierarchical clustering (according to the linkage metric). When this clustering corresponds to the clustering predicted by a hypothesis (as it does in this case) and the accompanying p-value is small, the evidence that the hypothesis is correct is much stronger than is provided by a p-value alone.

Another advantage of the method is that it is insensitive to properties that are word-specific, such as length, frequency of usage, age of acquisition, etc. This is achieved by removing the mean of the columns of the distance matrix, i.e. the baseline N400 response to each word. The remaining values only reflect the change in N400 response when a word is preceded by different prime words. Furthermore, since the average linkage algorithm determines the distance between two clusters by computing the ratio between the mean within-cluster distance and the mean distance to every other cluster, the word-pairs relevant to the computation always cover the complete set of words. This means that the dendrogram reflects effects that are caused by the interaction between a target word and every other word in the stimulus set, rather than effects that cause a certain word to have an intrinsically weak or strong N400 response. This leaves the experimenter with much more freedom in how to select the stimuli for the experiment.

The construction of a full word-to-word distance matrix of n items requires the presentation of $n^2 - n$ stimuli, hence the number of items that can be included in the analysis is restricted. Since the method can more reliably reveal patterns in semantic relationships when there are clearly distinguishable clusters in the stimulus set, the items that are included should be carefully chosen. Also, although the proposed method is unsupervised and will always produce some clustering solution, a careful experimental design is needed to ensure that the result is interpretable. This includes deciding at what level to "cut" the dendrogram.

In addition to answering a predefined research question, post-hoc analysis of the dendrogram may be used as a starting point for future exploration. For example, in our study, in addition to the top-level clusters, the dendrogram also hints at a dichotomy among the selected furniture stimuli. Indeed, strong semantic clusters may well exist within this category of words, for example based on the room that the furniture pieces are commonly found in. While the present study does not include

enough data to confirm such a hypothesis, the method suggests that this line of inquiry may be fruitful.

5 Conclusion

We have demonstrated a way to employ amplitude measurements of the N400 ERP component as a semantic distance metric between words. In order to obtain a reliable measurement, a multivariate analysis procedure based on the LCMV beamformer was successfully employed to overcome the low SNR of EEG signals. The resulting distance metric allows for successful application of unsupervised techniques, such as hierarchical clustering, to analyze how a chosen set of stimuli cluster together.

Our results illustrate how unsupervised techniques can be leveraged to analyze EEG data without strict adherence to predefined labels. This is particularly useful when validating theories concerning the organization of memory systems in the brain.

6 Acknowledgements

MvV was supported by the Interuniversity Attraction Poles Programme – Belgian Science Policy (IUAP P7/11) and is currently supported by a grant from the Aalto Brain Centre (ABC). MMVH is supported by research grants received from the Financing program (PFV/10/008), an interdisciplinary research project (IDO/12/007), and an industrial research fund project (IOF/HB/12/021) of the KU Leuven, the Belgian Fund for Scientific Research – Flanders (G088314N, G0A0914N), the Interuniversity Attraction Poles Programme – Belgian Science Policy (IUAP P7/11), the Flemish Regional Ministry of Education (Belgium) (GOA 10/019), and the Hercules Foundation (AKUL 043). RS is supported by the Academy of Finland (255349, 256459, 283071; LASTU programme 256887) and the Sigrid Jusélius Foundation.

References

Bates, D., Maechler, M., Bolker, B. M., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal Of Statistical Software, 67*(1), 1–48. doi:10.18637/jss.v067.i01

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision, 10*, 433–436. doi:10.1163/156856897X00357

Chan, A. M., Halgren, E., Marinkovic, K., & Cash, S. S. (2011). Decoding word and category-specific spatiotemporal representations from MEG and EEG. *NeuroImage, 54*(4), 3028–3039. doi:10.1016/j.neuroimage.2010.10.073

Collins, A. M. & Loftus, E. F. (1975). Spreading-activation theory of semantic memory. *Psychological Review, 82*(6), 407–428. doi:10.1037/0033-295X.82.6.407

Croft, R. J. & Barry, R. J. (2000). Removal of ocular artifact from the EEG: a review. *Neurophysiologie Clinique, 30*(1), 5–19. doi:10.1016/S0987-7053(00)00055-1

De Deyne, S., Navarro, D. J., & Storms, G. (2013). Better explanations of lexical and semantic cognition using networks derived from continued rather than single-word associations. *Behavior Research Methods, 45*(2), 480–498. doi:10.3758/s13428-012-0260-7

De Deyne, S., Navarro, D. J., & Perfors, A. (2016). Structure at every scale: a semantic network account of the similarities between unrelated concepts. *Journal of Experimental Psychology: General, 145*(7), IN PRESS. doi:10.1037/xge0000192

De Deyne, S. & Storms, G. (2008). Word associations: norms for 1,424 Dutch words in a continuous task. *Behavior Research Methods, 40*(1), 198–205. doi:10.3758/BRM.40.1.198

Debruille, J. B. & Renoult, L. (2009). Effects of semantic matching and of semantic category on reaction time and N400 that resist numerous repetitions. *Neuropsychologia, 47*(2), 506–517. doi:10.1016/j.neuropsychologia.2008.10.007

Friston, K. J., Stephan, K. M., Heather, J. D., Frith, C. D., Ioannides, A. A., Liu, L. C., … Frackowiak, R. S. (1996). A multivariate analysis of evoked responses in EEG and MEG data. *NeuroImage, 3*(3), 167–74. doi:10.1006/nimg.1996.0018

Gerlach, C. (2007). A review of functional imaging studies on category specificity. *Journal of Cognitive Neuroscience, 19*(2), 296–314. doi:10.1162/jocn.2007.19.2.296

Hunter, J. D. (2007). Matplotlib: a 2D graphics environment. *Computing in Science and Engineering, 9*(3), 99–104. doi:10.1109/MCSE.2007.55

Hutchison, K. A. (2003). Is semantic priming due to association strength or feature overlap? A microanalytic review. *Psychonomic Bulletin & Review, 10*(4), 785–813. doi:10.3758/BF03196544

Huth, A. G., De Heer, W. A., Griffiths, T. L., Theunissen, F. E., & Jack, L. (2016). Natural speech reveals the semantic maps that tile human cerebral cortex. *Nature, 532*(7600), 453–458. doi:10.1038/nature17637

Huth, A. G., Nishimoto, S., Vu, A. T., & Gallant, J. L. (2012). A continuous semantic space describes the representation of thousands of object and action categories across the human brain. *Neuron, 76*(6), 1210–1224. doi:10.1016/j.neuron.2012.10.014

Jain, A. K., Murty, M. N., & Flynn, P. J. (1999). Data clustering: a review. *ACM Computing Surveys, 31*(3), 264–323. doi:10.1145/331499.331504

Jones, M. N., Willits, J., & Dennis, S. (2015). Models of semantic memory. In *Oxford handbook of mathematical and computational psychology* (Chap. 11, pp. 232–254). Oxford University Press.

Koivisto, M. & Revonsuo, A. (2001). Cognitive representations underlying the N400 priming effect. *Cognitive Brain Research, 12*(3), 487–490. doi:10.1016/S0926-6410(01)00069-6

Kutas, M. & Federmeier, K. D. (2000, December). Electrophysiology reveals semantic memory use in language comprehension. *Trends in Cognitive Sciences, 4*(12), 463–470. doi:10.1016/S1364-6613(00)01560-6

Kutas, M. & Federmeier, K. D. (2011). Thirty years and counting: finding meaning in the N400 component of the event related brain potential (ERP). *Annual Review of Psychology, 62*, 621. doi:10.1146/annurev.psych.093008.131123

Kutas, M. & Hillyard, S. A. (1984). Brain potentials during reading reflect word expectancy and semantic association. *Nature, 307*(5947), 161–163. doi:10.1038/307161a0

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2015). LmerTest: Tests for random and fixed effects for linear mixed effect models. R package, version 2.0-29.

Luka, B. J. & Van Petten, C. (2014). Prospective and retrospective semantic processing: prediction, time, and relationship strength in event-related potentials. *Brain and Language, 135*, 115–129. doi:10.1016/j.bandl.2014.06.001

Martin, A. (2007). The representation of object concepts in the brain. *Annual Review of Psychology, 58*(1), 25–45. doi:10.1146/annurev.psych.57.102904.190143

McNamara, T. P. & Holbrook, J. B. (2003). Semantic memory and priming. In A. F. Healy & R. W. Proctor (Eds.), *Handbook of psychology: experimental psychology, vol. 4*. (pp. 447–474). New York: Wiley. doi:10.1002/0471264385.wei0416

McRae, K., Cree, G. S., Seidenberg, M. S., & McNorgan, C. (2005). Semantic feature production norms for a large set of living and nonliving things. *Behavior Research Methods*, *37*(4), 547–559. doi:10.3758/BRM.40.1.183

Neely, J. H. (1976). Semantic priming and retrieval from lexical memory: evidence for facilitatory and inhibitory processes. *Memory & Cognition*, *4*(5), 648–654. doi:10.3758/BF03213230

Neely, J. H. (1991). Semantic priming effects in visual word recognition: a selective review of current findings and theories. In D. Besner & G. W. Humphreys (Eds.), *Basic processes in visual word recognition* (Chap. 9, pp. 264–323). Hillsdale, NJ: Lawrence Erlbaum Associates.

Nelson, D. L., McEvoy, C. L., & Schreiber, T. A. (2004). The University of South Florida free association, rhyme, and word fragment norms. *Behavior Research Methods, Instruments, & Computers*, *36*(3), 402–407. doi:10.3758/BF03195588

Norman, K. A., Polyn, S. M., Detre, G. J., & Haxby, J. V. (2006). Beyond mind-reading: multi-voxel pattern analysis of fMRI data. *Trends in Cognitive Sciences*, *10*(9), 424–430. doi:10.1016/j.tics.2006.07.005

Oliphant, T. E. (2007). Python for scientific computing. *Computing in Science and Engineering*, *9*(3), 10–20. doi:10.1109/MCSE.2007.58

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... Duchesnay, É. (2012). Scikit-learn: machine learning in python. *Journal of Machine Learning Research*, *12*, 2825–2830. doi:10.1007/s13398-014-0173-7.2

Perrin, F., Pernier, J., Bertrand, O., & Echallier, J. F. (1989). Spherical splines for scalp potential and current density mapping. *Electroencephalography and Clinical Neurophysiology*, *72*(2), 184–187. doi:10.1016/0013-4694(89)90180-6

Pulvermüller, F. (2013). How neurons make meaning: brain mechanisms for embodied and abstract-symbolic semantics. *Trends in Cognitive Sciences*, *17*(9), 458–470. doi:10.1016/j.tics.2013.06.004

R Core Development Team. (2015). *R: a language and environment for statistical computing, 3.2.1*. R Foundation for Statistical Computing. Vienna, Austria. doi:10.1017/CBO9781107415324.004

Renoult, L. & Debruille, J. B. (2011). N400-like potentials and reaction times index semantic relations between highly repeated individual words. *Journal of Cognitive Neuroscience*, *23*(4), 905–922. doi:10.1162/jocn.2009.21410

Renoult, L., Wang, X., Calcagno, V., Prévost, M., & Debruille, J. B. (2012). From N400 to N300: variations in the timing of semantic processing with repetition. *NeuroImage*, *61*(1), 206–215. doi:10.1016/j.neuroimage.2012.02.069

Rips, L. J., Shoben, E. J., & Smith, E. E. (1973). Semantic distance and the verification of semantic relations. *Journal of Verbal Learning and Verbal Behavior*, *12*(1), 1–20. doi:10.1016/S0022-5371(73)80056-8

Rugg, M. D. & Curran, T. (2007). Event-related potentials and recognition memory. *Trends in Cognitive Sciences*, *11*(6), 251–257. doi:10.1016/j.tics.2007.04.004

Satterthwaite, F. E. (1946). An approximate distribution of estimates of variance components. *Biometrics*, *2*(6), 110–114. doi:10.2307/3002019

Simanova, I., van Gerven, M., Oostenveld, R., & Hagoort, P. (2010). Identifying object categories from event-related EEG: toward decoding of conceptual representations. *PLoS ONE*, *5*(12), E14465. doi:10.1371/journal.pone.0014465

Smith, N. J. & Kutas, M. (2015). Regression-based estimation of ERP waveforms: I. The rERP framework. *Psychophysiology*, *52*(2), 157–168. doi:10.1111/psyp.12317

Sokal, R. R. & Michener, C. D. (1958). A statistical method for evaluating systematic relationships. *University of Kansas Science Bulletin*, *38*(22), 1409–1438. doi:citeulike-article-id:1327877

Treder, M. S., Porbadnigk, A. K., Shahbazi Avarvand, F., Müller, K.-R., & Blankertz, B. (2016). The LDA beamformer: Optimal estimation of ERP source time series using linear discriminant analysis. *NeuroImage*, *129*, 279–291. doi:10.1016/j.neuroimage.2016.01.019

Van Petten, C. (2014). Examining the N400 semantic context effect item-by-item: relationship to corpus-based measures of word co-occurrence. *International Journal of Psychophysiology*, *94*(3), 407–419. doi:10.1016/j.ijpsycho.2014.10.012

Van Petten, C. K. (1993). A comparison of lexical and sentence-level context effects in event-related potentials. *Language and Cognitive Processes1*, *8*(4), 485–531.

Van Strien, J. W., Hagenbeek, R. E., Stam, C. J., Rombouts, S. A. R. B., & Barkhof, F. (2005). Changes in brain electrical activity during extended continuous word recognition. *NeuroImage*, *26*(3), 952–959. doi:10.1016/j.neuroimage.2005.03.003

Van Veen, B. D., Van Drongelen, W., Yuchtman, M., & Suzuki, A. (1997). Localization of brain electrical activity via linearly constrained minimum variance spatial filtering. *IEEE Transactions on Biomedical Engineering*, *44*(9), 867–880. doi:10.1109/10.623056

van Vliet, M., Chumerin, N., De Deyne, S., Wiersema, J. R., Fias, W., Storms, G., & Van Hulle, M. M. (2016). Single-trial ERP component analysis using a spatiotemporal LCMV beamformer. *IEEE Transactions on Biomedical Engineering*, *63*(1), 55–66. doi:10.1109/TBME.2015.2468588

van Vliet, M., Manyakov, N. V., Storms, G., Fias, W., Wiersema, J. R., & Van Hulle, M. M. (2014). Response-related potentials during semantic priming: the effect of a speeded button response task on ERPs. *PLoS ONE*, *9*(2), E87650. doi:10.1371/journal.pone.0087650

Wittevrongel, B. & Van Hulle, M. M. (2016, May). Faster P300 classifier training using spatiotemporal beamforming. *International Journal of Neural Systems*, *26*(03), 1650014. doi:10.1142/S0129065716500143