Very early and late form-to-meaning computations during visual word recognition as revealed by electrophysiology

Simone Sulpizio[1], Giorgio Arcara[2], Sara Lago[2,4], Marco Marelli[1] & Simona Amenta[3],

[1]University of Milano-Bicocca, Milan, Italy

[2]IRCCS San Camillo Hospital, Venice, Italy

[3]University of Trento, Trento, Italy

[4] Padova Neuroscience Center, University of Padova, Italy

Running Head: FORM-TO-MEANING COMPUTATIONS DURING VISUAL WORD RECOGNITION

Corresponding author:

Giorgio Arcara

IRCCS San Camillo Hospital, Venice, Italy

Via Alberoni, 70 30126 – Lido, Venezia (VE), Italy

email: giorgio.arcara@gmail.com

Author Contributions: SS, MM, GA, and SA designed the study. GA and SL analyzed the data. SS, MM, SS and SA gave the main interpretations to the results. SS drafted the first version of the paper, with contributions from SA and GA. All authors commented and actively contributed to produce the final version of the paper.

Declarations of interest: none

Abstract

We used a large-scale data-driven approach to investigate the role of word form in accessing semantics. By using distributional semantic methods and taking advantage of an ERP lexical decision mega-study, we investigated the exact time dynamic of semantic access from printed words as driven by orthography-semantics consistency (OSC) and phonology-semantics consistency (PSC). Generalized Additive Models revealed very early and late OSC-by-PSC interactions, visible at 100 and 400 ms, respectively. This pattern suggests that, during visual word recognition: a) meaning is accessed by means of two distinct and interactive paths – the orthography-to-meaning and the orthography-to-phonology-to-meaning path –, which mutually contribute to recognition since early stages; b) the system may exploit a dual mechanism for semantic access, with early and late effects associated to a fast-coarse and a slow-fine grained semantic analysis, respectively. The results also highlight the high sensitivity of the visual word recognition system to arbitrary form-meaning relations.

Keywords: phonology-semantics consistency, orthography-semantics consistency, EEG, visual word recognition, form-meaning relation.

1. Introduction

Reading is the unique human ability to use visual symbols to access the meaning and sound of words. Although learning to read requires years, once acquired, this ability is performed quasi-automatically and near-instantaneously: When facing a written string, the brain of a skilled adult reader can extract semantic, orthographic, and phonological information in less than half a second. Although there is consensus that, during reading, people access these three types of information – i.e., orthography, phonology, and semantics –, the time dynamic of this access is still controversial. At the bulk of the controversy there is the question as to whether higher-level information (i.e., word meaning) can contribute to the processing of lower-level information (i.e., word form). The aim of the present study is to answer this question, by investigating the time-course of the interplay between form and meaning in reading.

From a neurocognitive perspective, a widely accepted view of reading assumes that during word recognition, the extraction of visuo-orthographic features occurring within the left ventral occipito-temporal cortex is the door for accessing and processing phonological and semantic information, which comes into play only (relatively) late in time. Support for such a perspective, not only comes from experiments using metabolic neuroimaging methods and assuming a feed-forward processing dynamic of the word recognition process – i.e., information flow strictly proceeds from low-level visual features to word representations (e.g., Dehaene & Cohen, 2011; Maurer et al., 2011; Perrone-Bertolotti et al., 2017; Schurz et al., 2014; Vinckier et al., 2007; but note that these methods have poor time resolution and are thus inadequate to answer questions about the time dynamics of brain processing) –, but also from studies using electrophysiological methods, which usually identify an early and a late stage of processing during word recognition. In particular, the early stage occurs within the first ~300 ms after stimulus presentation and is mainly associated to visuo-orthographic processing and orthography-to-phonology mapping, whereas the late stage goes from ~300 ms onwards and is associated with lexical-semantic processing (e.g., Bentin et al., 1999; Carrasco-Ortiz et al., 2017; Grainger & Holcomb, 2009; Stites & Laszlo, 2017). Recently, Laszlo and Federmeier (2014) recorded ERPs in a lexical decision experiment with the aim to investigate when orthographic, lexical, and semantic variables affected word recognition. Their single-level item analyses showed that orthographic effects started to be reliable after ~130ms, word frequency effects after ~270ms, and semantic effects only after ~300ms. A similar time dynamic has been also reported by Dufau et al. (2015), who ran an ERP lexical-decision mega-study and found that the early ~300 ms were affected by orthographic and lexical variables (e.g., word length, orthographic Levenshtein distance), but not by semantic ones (i.e., concreteness). The authors concluded that their results suggest "a fast initial feed-forward sweep of neural activity cascading through visual, orthographic, and lexical representations" (p. 1895).

An alternative view, however, has been also advanced, in which reading is conceptualized as an extremely interactive process, with all the levels of analysis – i.e., orthography, phonology, semantics – synergistically working since the early stage of processing (Harm & Seidenberg, 2004; McClelland, 1979; for a neurocognitive proposal, see Price & Devlin, 2011). In such a view, thanks to the strong learned associations between semantics and orthography, semantic information may become available nearly simultaneously with orthographic processing (e.g., Pulvermüller et al., 2009) and actively contribute to orthographic decoding. Some evidence in accordance with this view has also emerged from a few EEG studies reporting signs of early semantic processing and thus challenging the processing unfolding typically advocated by the strict feed-forward perspective (e.g., Chan et al., 2011). Hauk et al. (2012) recorded ERPs while participants performed a lexical decision and a semantic categorization task. In both tasks, lexical and semantic information were activated nearly simultaneously, starting to show an effect ~160 ms after word onset (see also Hauk, Patterson, et al. 2006b). Amsel et al. (2013) measured ERPs during a go/no-go semantic categorization task and found that information differentiating living from nonliving things becomes available in 160 ms. Since the same result was not found for information differentiating graspable from ungraspable objects, the authors suggested that the initial semantic processing computes only coarse-grained conceptual information. A similar conclusion was also drawn by Louwerse and Hutchinson (2012), who used a semantic judgment task (and an iconicity task) to investigate the time-course of conceptual processing in relation to grounded simulation and statistical linguistic frequencies. The authors found that conceptual-linguistic processing precedes conceptual-grounded processing, starting within 100 ms from target onset, and concluded that the system quickly extracts meaning through language statistics in order to provide a first rough representation, which is then qualified by the subsequent (slower) grounded simulation.

Early effects of semantics have been reported also by Chen et al. (2015), who used an EEG/MEG recording to assess how the occipitotemporal cortex responds to orthographic, lexical and semantic variables in silent reading, lexical decision, and semantic categorization. Their results showed task-dependent semantic effects around 160 ms in an occipitotemporal region that was also found to be highly sensitive to word frequency: This pattern speaks in favor of an interactive view, in which all sources of information jointly contribute to optimize word recognition. Note, however, that the appearance of early interactive effects seems to be (at least in part) task-dependent – early interactivity is found with semantic categorization and reading, but not with lexical decision. Thus, early interactivity might be more a possibility the system exploits when the task requires it, rather than an intrinsic property of the recognition process.

In the current study, we present data on the time dynamic of visual word recognition (in a lexical decision task) and provide clear-cut evidence that semantic processing affects recognition since its earliest stage, in an early interaction between conceptual information and form-related information. We used new methods both to analyze ERP data and to measure the interaction between semantic (i.e., conceptual) information on the one hand, and orthographic and phonological (i.e., form-related) information on the other hand. Regarding the latter, our approach is grounded in methods from distributional semantics (Landauer & Dumais, 1997; Turney & Pantel, 2010), that can provide cognitively-sound estimates of semantic representations (e.g., Gunther, Rinaldi & Marelli, 2019). In Distributional Semantics Models (DSM) the meaning of a word can be estimated by the way that it co-occurs with other words in the whole lexicon. Word meanings are represented as vectors based on these co-occurrences: The more two words occur in similar contexts (i.e., flanked by comparable sets of words), the more their vectors will be close and their meanings are similar (and *vice versa*). Similarity is geometrically expressed as the cosine of the angle formed between two vectors: The more similar the two vectors, the smaller the angle, the higher their cosine.

The DSM approach has been successfully adopted in word recognition studies aimed at analyzing the interplay of form and semantics in visual word recognition (Marelli, Amenta & Crepaldi, 2015; Amenta, Marelli & Sulpizio, 2017; Marelli & Amenta, 2018; Amenta, Crepaldi & Marelli, 2020; Siegelman et al., 2022). In these studies, DSMs were used to develop new measures that quantify the relationship between form and meaning, i.e., Orthography- and Phonology-Semantics Consistency (OSC and PSC). OSC is an estimate of semantic similarity between a string of letters (e.g., widow) and all its orthographic relatives – i.e., the words in the lexicon that embed that same sequence (e.g., widower, widowhood, etc.). Mathematically, OSC is formalized as the frequency-weighted average semantic similarity between the vector of a target word and the vectors of all the words that contain it (see below). This estimate tells us how consistent is the mapping between form and meaning for each word. Words with high OSC are those whose orthographic relatives are also semantically associated (e.g., widow, widower, widowhood, etc.). PSC is the phonological counterpart to OSC: it is based on the same methods of OSC, but takes into account phonological relatives in its formalization. In an information-theoretic perspective, OSC and PSC are measures of the degree of uncertainty in meaning access, as respectively informed by orthography and phonology. That is, these metrics capture to what extent orthographic and phonological features of a word are ambiguous in cueing its meaning: low OSC and PSC scores indicate that the orthography and phonology of a given word are associated to a wide, inconsistent semantic range.

Although computed from word forms, OSC and PSC are measures that allow to test the effective contribution of form-meaning association to word recognition without the need of explicit semantic tasks or manipulations. Therefore, these measures are good candidates to investigate the alleged early interactivity of the recognition process. Both OSC and PSC proved to be relevant predictors of lexical decision latencies in visual word recognition (over and above well known lexical and semantic predictors, among which: word frequency, length, neighbourhood size, etc; see Marelli et al., 2015; Amenta et al., 2017; Marelli & Amenta, 2018), showing that, during word recognition, semantics is accessed by a mutual interaction between phonological and orthographic information (Amenta et al., 2017). However, what is still unclear from these results is the time-course of this interaction. Specifically, we do not know how early semantic information starts affecting the process. In fact, previous studies only analyzed behavioural responses, which are silent on the time dynamics between phonological, orthographic and semantic information *during* word processing. As a result, we do not know when form-to-meaning consistencies begin to exert their influence during word recognition. In the present study, for the first time, we investigated the impact of OSC and PSC on ERPs, which are excellent measures to investigate the temporal unfolding of cognitive processing. In doing so, we were able to isolate the time course of semantic processing (as informed by orthography and phonology) and identify, with precision, whether it starts playing a role during the very early stage of word recognition or only later on. With respect to our purpose, OSC and PSC are of great interest for multiple reasons: 1) They effectively capture semantic aspects, since they quantify the mapping between word form and word meaning; 2) they are quantitatively and automatically determined, and give back quantitative and easy-to-interpret information; 3) they are mainly a-theoretical, with no need to postulate a-priori, researcher-defined semantic properties; 4) most importantly, they are based on a neurally plausible measure of semantics (Sassenhagen & Fiebach, 2020; Just et al., 2010; Mitchell et al., 2008; Pereira et al., 2018) and may contribute to the development of theories of neural representations. To answer our research question, we capitalized on the ERP mega-study by Dufau et al. (2015) who collected data on ~1,000 words from 75 participants performing a very simple task, i.e., lexical decision. This dataset is ideally suited to investigate potential effects of OSC and PSC. In fact, the recent analyses by Sassenhagen and Fiebach (2020), using the same dataset, showed that these ERP data encode word properties captured by distributional semantic models, the very same models at the foundation of OSC and PSC. To detect potential early semantic effects, and to take into account the potentially complex interactions between variables, we analyzed ERP data by fitting tensor multi-dimensional surfaces with Generalized Additive Models (GAMs; Tremblay & Newman, 2015; Wood, 2017). Offering a convenient way to model complex interactions between continuous variables (Kuperman et al., 1995), this state-of-the-art approach is particularly suitable for the present study, whose main aim is to model the exact time dynamics of the interaction between semantic, orthographic and phonological information (i.e., investigating the unfolding of the three-way interaction between Time, OSC, and PSC).

2. Methods

2.1 Resource and stimuli.

Stimuli were extracted from the Dufau et al.'s (2015) mega-study, which contains item-level ERP data for 960 words (with at least 43 trials per word) recorded from 29 sites on the scalp (FP1, FPz, FP2, F7, F3, Fz, F4, F8, FC5, FC1, FC2, FC6, C3, Cz, C4, T7, T8, CP5, CP1, CP2, CP6, P7, P3, Pz, P4, P8, O1, Oz, O2). Reference was placed on the left mastoid. Data were collected from 75 right-handed healthy young adults (age range: 18-25), all English native speakers. Participants performed a go/no-go lexical decision task in which they were asked to press a button as soon as possible whenever they detected a nonword. Each trial started with a 400-ms presentation of a letter string followed by a 600-ms black screen. Data are freely available in preprocessed (i.e., filtered, artifact free, and baseline corrected) epochs going from 100 ms before to 920 ms after stimulus presentation (for further methodological details on data collection and preprocessing, see Dufau et al., 2015).

We selected a set of 689 English words for which it was possible to compute separate estimates of OSC and PSC that dissociated one from the other (see Amenta et al., 2017), for which either PSC or OSC was different from 1 (see Marelli & Amenta, 2018, for a discussion), and that were also included in the Kilo-word database (Dufau et al., 2015). Table 1 reports descriptive data for the experimental set. Frequency values were extracted from SUBTLEX-UK (van Heuven, Mandera, Keuleers, & Brysbaert, 2014).

	Mean	SD	Q1	Median	Q3
Log Frequency	2.63	0.71	2,24	2,77	3.15
Orthographic Length	5.68	1.38	4	6	7
OSC	0.79	0.25	0.69	0.89	0.98
PSC	0.76	0.28	0.65	0.89	0.98

Table 1. Descriptive data for the experimental items

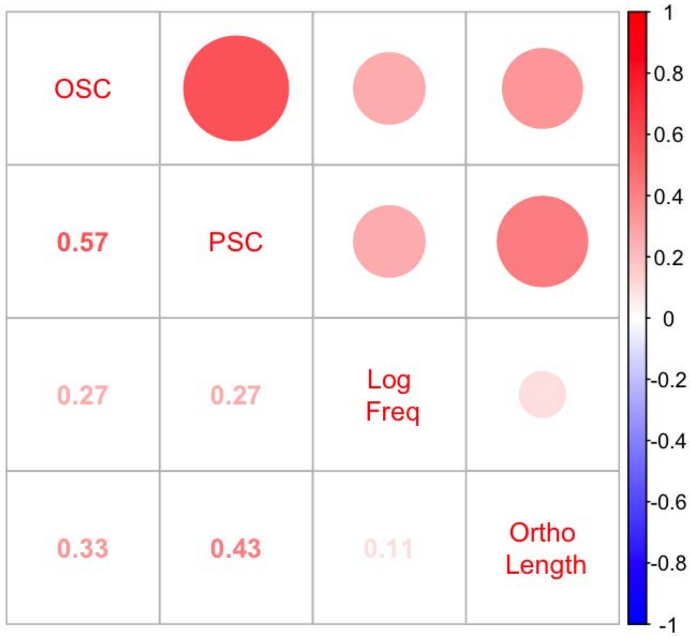


Figure 1. Correlations among psycholinguistic variables of the experimental stimuli. The figure reports the Pearson correlation among variables of the stimuli. Larger circles indicate higher correlation indexes. Warm and cold colors indicate positive and negative correlations, respectively.

2.2 Measures.

OSC and PSC were computed following the same procedure and sources described in Amenta et al. (2017). OSC is defined as:

$$OSC(t)=\frac{\sum_{x=1}^{k}\cos\left(\vec{t},\vec{r}_x\right)*f_{r_x}}{\sum_{x=1}^{k}f_{r_x}}$$

Where *t* is the target word, r_x each of its *k* orthographic relatives, and f_{rx} the corresponding frequencies. Following Marelli and Amenta (2018), we considered as orthographic relatives each word containing the target (e.g., unicorn, cornfield, corner, scornful, etc., were all relatives of corn) in the 65,000-word list identified by Marelli and Amenta (2018). In order to quantitatively capture word meanings, represented by $\vec{t}$ and $\vec{r_x}$ in the above formula, we relied on distributional semantics. This computational approach builds on lexical co-occurrences to induce meaning representations in the form of vectors, whose proximity can then be used as a proxy for semantic relatedness. Distributional semantics is a popular approach in the modelling of semantic memory, with several proposals advanced in the cognitive science literature (for recent reviews see Günther, Rinaldi, & Marelli, 2019; Jones, Willits, & Dennis, 2015).

A concatenation of UkWac, Wikipedia, and BNC served as the base to build the distributional model, trained using the word2vec tool (Mikolov, Sutskever, Chen, Corrado, & Dean, 2013). Model parameters were selected following Baroni, Dinu and Kruszewsky (2014): CBOW method with 400-dimension vectors, considering a 5-word window (see Marelli & Amenta (2018) for further details on computation and validation of the obtained measures). We used the obtained model to compute cos(t,s), that is, the semantic similarity between a target vector and the vectors of each of its relatives.

The same procedure and formalization above were used to obtain PSC. The only difference in the computation lies in how the relatives are defined. For PSC, relatives are phonologically defined: A phonological relative is a word whose phonological form contains the phonological sequence of the target word (e.g., basin /ˈbeɪsən/, bacon /ˈbeɪkən/, debate /dɪˈbeɪt/, etc., were all relatives of bay /beɪ/). Phonological annotations were extracted from CELEX (Baayen et al., 1995).

For more details on how OSC and PSC were computed, we refer to Marelli & Amenta (2018), and Amenta et al. (2017).

2.3 ERP Statistical analyses.

To assess the effects of OSC and PSC over time we used Genaralized Additive Models (GAMs, Tremblay & Newman, 2015; Wood, 2017). GAMs are an extension of general linear models (GLMs), that can easily handle non-linear relationship between the predictors and the dependent variable, and that in recent years has been successfully applied to ERP studies of word processing (e.g., De Cat et al., 2015; Hendrix et al., 2016; Kryuchkova et al., 2012). Although it is possible to fit non-linearities with GLMs (for example by including polynomial terms), this should be done with a-priori choices and only a relatively limited number of non-linear relations can be modeled. GAMs allow to overcome these limitations: in particular, the non-linear relationship between predictors and the dependent variable are modeled in a bottom-up fashion with smoothing splines. The actual number of splines used to model the relationship between predictors and dependent variables and the parameters of these splines are determined in a bottom-up fashion, according to some criteria. A main advantage of GAMs, as compared to traditional GLMs, is that they also allow a convenient way to model complex interactions between continuous variables (Baayen, Kuperman, & Bertram, 2010). Shortly, fitting interactions between continuous variables is potentially challenging for linear models, because the effects are bound to some specific constraints (of the imposed linear trends) that may cause misfits in the data, especially for high or low values of the predictors for which the leverage is high. This is due to the relying on a multiplicative approach (typically applied to interactions used in multiple regression models), that imposes a very specific functional form to the interaction (Baayen, Kuperman, & Bertram, 2010). GAMs, on the contrary, allow many possible ways to model interactions, and are hence better suited at capturing complex dynamics in the data. This is particularly relevant for the present study, in which our main aim is to model a three-way interaction between Time, OSC, and PSC, that is how the interaction between OSC, and PSC changes over time.

In the present implementation, parameters are fit according to the Maximum Likelihood method (Wood, 2017).

We fitted a series of separate GAMs (one for each electrode) with the following syntax[1]:

[1] Please note, that this is a simplified syntax. See the Supplementary Materials for full R code used for the analysis.

```
(1)     Ampl~
        # main effects
        s(WordFrequency) + s(Time) +
        s(NumberOfLetters) + s(psc) + s(osc) +
        # two-way interactions with Time
        ti(WordFrequency, Time, k=c(3,10))+
        ti(NumberOfLetters, Time, k=c(3,10))+
        ti(psc, Time) +
        ti(osc, Time) +
        # two-way interaction of psc and osc
        ti(psc, osc) +
        # three-way interaction of psc, osc, time
        ti(psc,osc,Time, k=c(3,3,10)) +
        # random intercepts
        s(WORD, bs="re"))
```

The syntax above indicates that, in each model, Amplitude was the dependent variable, explained by several predictors: an interaction between Word Frequency and Time, an interaction between Number of Letters and Time, and an interaction between PSC, OSC, and Time (note that the order of terms in the interaction is irrelevant).

All interactions were modeled by means of tensors (i.e., a function that allows modeling interactions between continuous variables). The parameters specified as “k” indicate the number of basis functions that are used to define the tensors (Linke & Baayen, 2019). To limit the overfitting of the data, we opted for a relatively low number of basis functions for Word Frequency, Number of Letters, PSC, and OSC (i.e., k=3), but we allowed for a relatively higher number for Time (k=10), in order to be able to capture the expected fluctuation over time of ERP amplitude. The term (WORD, bs="re") indicates a random effect for WORD in the syntax of the mgcv package. Note that as the Kilo-word corpus includes data averaged across participants we could not include participants as random factor. Finally, the correlation term refers to that autocorrelation between timepoints, which is highly expected in the case of time series and is hence taken into account in the model.

After fitting each model a series of diagnostic checks were performed. We inspected residual distribution and whether the relationship between fitted and observed values was satisfactory. Correlation among variables is not an issue for GAM, but concurvity (which is the GAM equivalent for collinearity) may be, and so we inspected concurvity for each term. Finally, as data consisted of time series and some autocorrelation was expected, we also took into account this parameter. The choice of autocorrelation parameters (rho = 0.1) was made after a preliminary fit of GAMs and inspecting the autocorrelation function of model residuals (ACF). Details can be found in Supplementary Materials.

Although it is generally expected that signals from EEG electrodes would be correlated (due to volume conduction of EEG sources), we opted not to explicitly model the dependency on electrode data, to avoid having a model with too many parameters. Given this choice of modeling separately data for each electrode, the qualitative inspection of the results on electrodes could be taken as diagnostic for model overfitting (De Cat et al., 2015). Figures were obtained using the erpR package (Arcara & Petrova, 2017) and custom code. The full code employed in the analysis is available online in the Open Science Framework (https://osf.io/4e7tq/). It is also possible to fully explore the results interactively via a ShinyApp available at this link https://giorgioarcara.shinyapps.io/ERP-OSC-PSC/.

No part of the study analyses was pre-registered prior to the research being conducted.

3. Results

Results on all models showed significant effects of the interaction between OSC, PSC, and Time for all electrodes (*ps*< 0.05). Full results for all the other terms are reported in the Appendix (Table A1 and A2). Overall, models showed a good fit with an explained variance ranging from 40% to 70%.

Thanks to the introduction of the term “s(Time)” as main effect in the models’ syntax, we were able to link up the effects of OSC and PSC to the typical average waveforms studied extensively in the ERP literature (see Figure 2), especially how the three-way interaction of time, PSC and OSC modulates this curve. These models, therefore, allow to uniquely characterize the in-time unfolding of the effects with extreme precision. For these reasons, when discussing our results we focus mainly on the timing of the effects and avoid to label them in terms of components (although we discuss them in relation to the available ERP literature). The main way to interpret results of GAMs with tensor surfaces (used to investigate the effect of interactions) is through visual inspection.

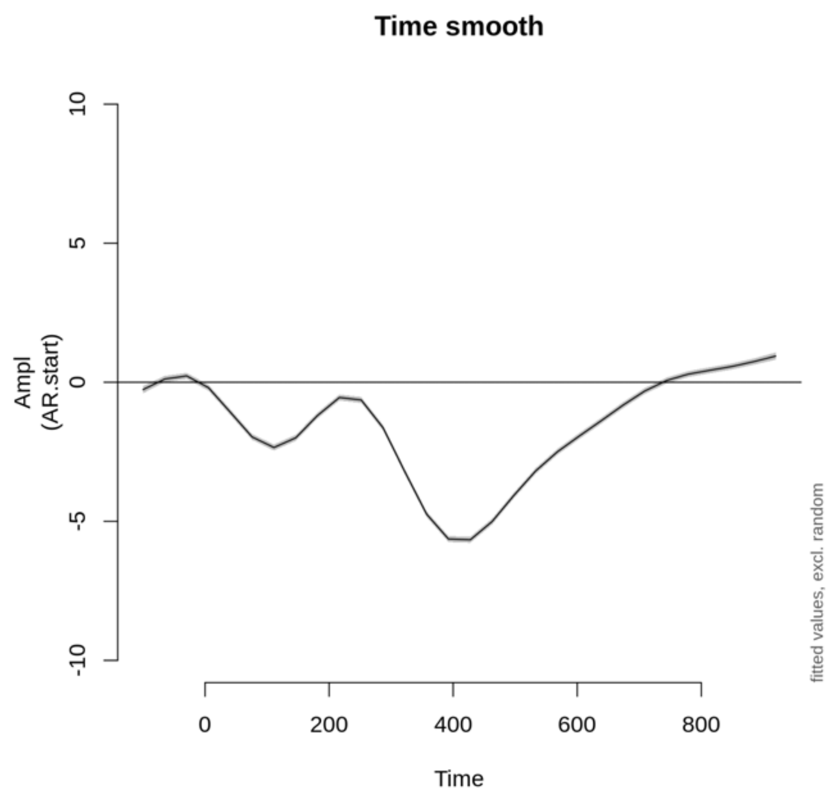


Figure 2. Time smooth for a representative electrode (FC1). The traditional positivities and negativities of ERP deflections are reflected in this smooth.

Before discussing the results for PSC and OSC, we inspected the pattern of results for Length and Frequency, which have been widely investigated in the ERP literature and can be thus considered as benchmark effects. The early effects of word length surfaced ~200 ms after stimulus onset, with more positive amplitude associated with shorter words and negative amplitudes to longer ones on centro-posterior sites (see Fig. A2 in Appendix). The effect is in line (despite slightly later) with previous findings (e.g., Assadollahi & Pulvermüller, 2003; Dufau et al., 2015; Hauk, Davis et al., 2006; Hauk & Pulvermüller, 2004). The effect of word frequency surfaced at ~300 ms and then at ~800 ms on fronto-central electrodes (see Fig. A3 in Appendix). The time dynamics of the effect is compatible with the pattern reported in the literature (e.g., e.g., Assadollahi & Pulvermüller, 2003; Dufau et al., 2015; Hauk, Davis et al., 2006; Hauk & Pulvermüller, 2004). Taken together, the results for the effects of length and frequency indicate the robustness and reliability of our findings. Further details and figures concerning these effects can be found in the Supplementary materials.

Turning to the variables of main interest (OSC and PSC), basing on our aims, we focused on two a-priori selected time points, that are 100 ms and 400 ms: while the former is the lower boundary at which semantic activation from printed words has been occasionally reported (Dell'Acqua et al., 2010, Louwerse & Hutchinson, 2012), the latter is the time range at which semantic effects are most typically observed (e.g., Dufau et al., 2015; Grainger & Holcomb, 2009; Lau et al., 2008). Figure 3 and 4 show the results of the interaction of OSC, PSC, and Time, at these two timepoints, 100 ms and 400 ms (for a plot of the grand average of ERPs for all electrodes from the Kilo-word dataset, see Figure A1 in Appendix; for inspecting the full time course of the interaction, see https://giorgioarcara.shinyapps.io/ERP-OSC-PSC). Neighbouring electrodes (for which models were calculated separately) showed similar results, as expected due to electrode amplitude similarity related to volume conduction.

As visible in Figure 3, a significant PSC x OSC interaction surfaces at 100 ms with a wide fronto-central distribution: At very low level of OSC, middle-to-high levels of PSC were associated with more negative amplitude. As exemplified in Figure 5 (lower panel, blue blobs), this effect was long lasting, being visible up to ~250 ms after target onset (for a point-by-point time course, see https://giorgioarcara.shinyapps.io/ERP-OSC-PSC). Figure 5 also shows that this early effect was absent at medium (0.5) and high level (0.75) of OSC (medium and upper panel, respectively).

A second effect also surfaced later on time, when we explored the second time interval of interest. Figure 4 shows this effect at 400 ms: in fronto-central electrodes, at low and intermediate levels of OSC, middle-to-high levels of PSC were associated with more negative amplitude. This effect was visible between ~300 and ~400 ms after word presentation (see also Figure 5 and 6, for details of the time course on specific electrodes; for a point-by-point time course on the whole scalp, see https://giorgioarcara.shinyapps.io/ERP-OSC-PSC). Finally, we note that, when exploring late processes (i.e., later than 400 ms after target presentation), a further effect surfaces between ~750 and ~900 ms, with a fronto-central distribution: At low levels of OSC, middle-to-high levels of PSC were associated with more positive amplitude. However, because of our hypotheses, our analyses were focused on two specific time points clearly associated with semantic processing. Thus, we do not further discuss unexpected findings raised by a-posteriori visual inspection of our results.

Diagnostics associated to the models showed also good fit to the data, normal distribution of residuals, good properties in terms of autocorrelation and a reasonable concurvity (i.e., the GAM equivalent of collinearity) for the main effect of interest (namely the three-way interaction between OSC, PSC and Time). A high concurvity was found for some nuance variables (that is Frequency and Length). Details on the diagnostics can be found in the osf link associated with the article (https://osf.io/4e7tq/).

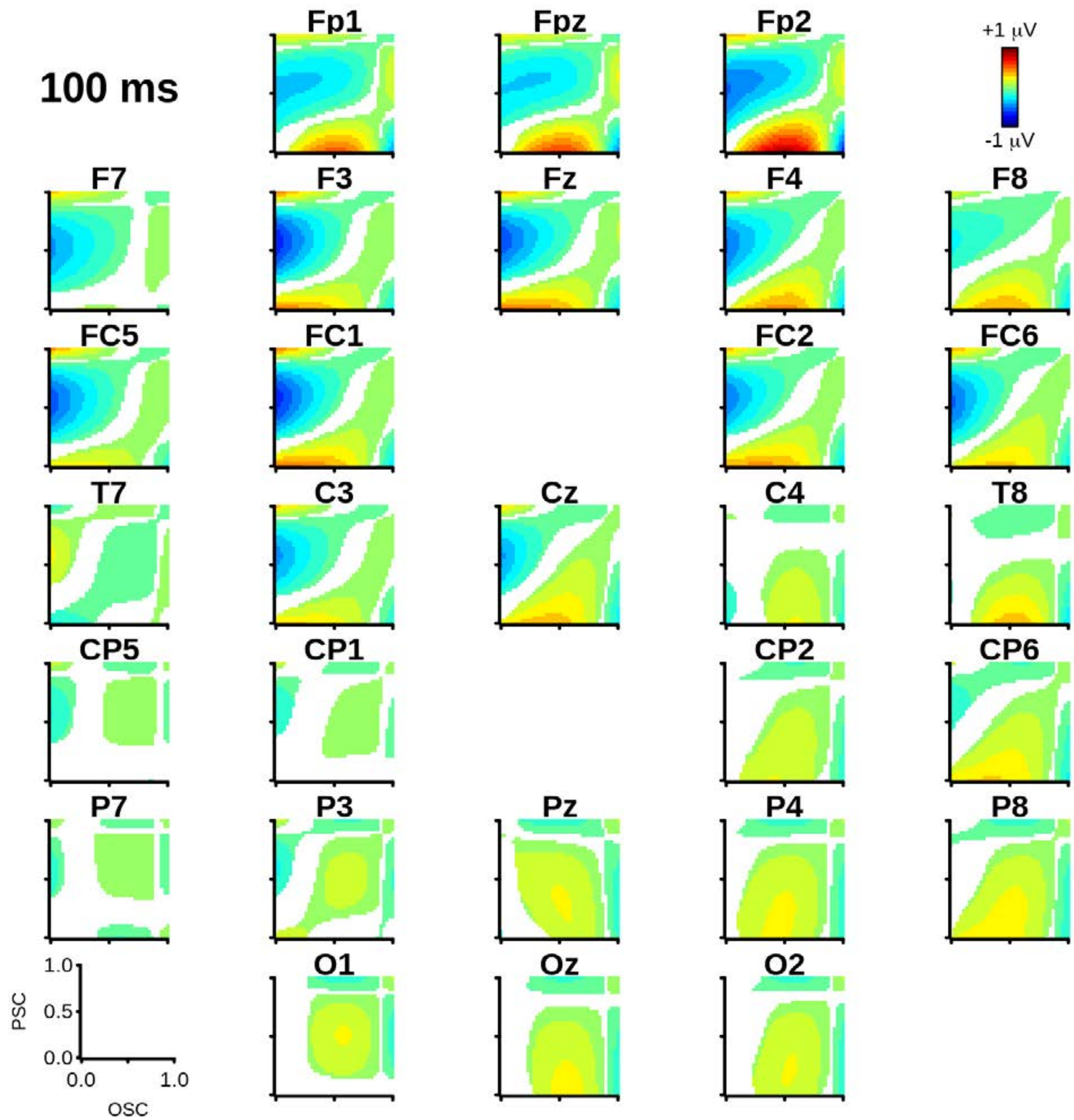


Figure 3. Scalp plot of partial effects of the interaction between OSC, PSC, and Time at 100 ms. Partial effect of OSC, PSC, and Time for each electrode at 100 ms. The contour map for each electrode is a topographic representation of the partial effect of the interaction between OSC (on the x-axis), PSC(on the-y axis), in a specific timepoint. Amplitude is codified as colours, using a jet palette: color towards red indicates positive values, while color towards blue indicates negative values, while colors

toward green indicate in-between values. Topographic maps have been masked so that only effect estimates whose confidence interval at 95% did not include zero were included. Electrodes are reported in a regular grid array that approximate their position on the scalp. Effect on all timepoints can be inspected with this app https://giorgioarcara.shinyapps.io/ERP-OSC-PSC

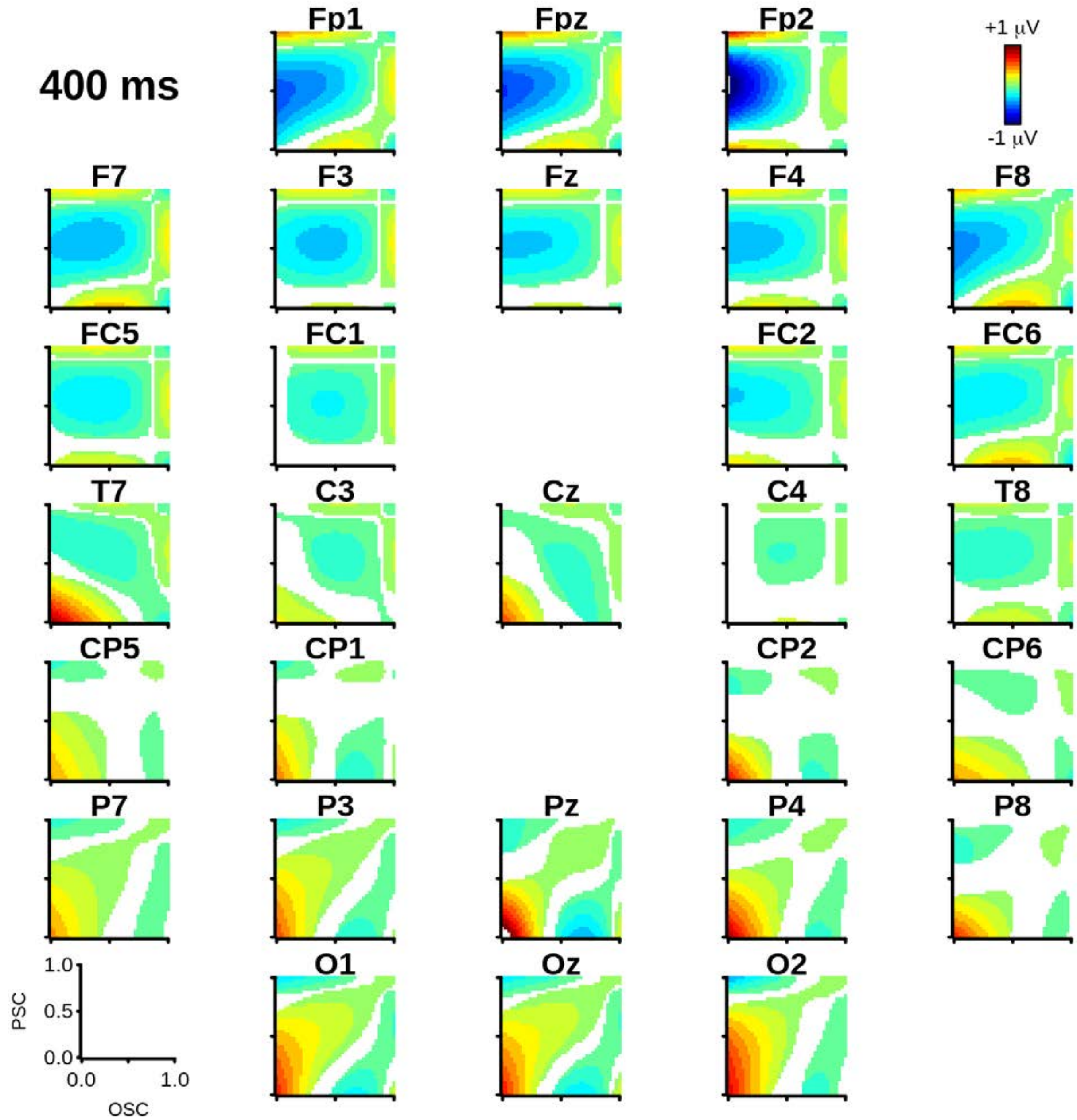


Figure 4. Scalp plot of partial effects of the interaction between OSC, PSC, and Time at 400 ms

after stimulus offset. The figure shows the partial effect of OSC, PSC, and Time for each electrode at 400 ms. The contour map for each electrode is a topographic representation of the partial effect of the interaction between OSC (on the x-axis), PSC(on the-y axis), in a specific timepoint. Amplitude is codified as colours, using a jet palette: color towards red indicates positive values, while color towards blue indicates negative values, while colors toward green indicate in-between values. Topographic maps have been masked so that only effect estimates whose confidence interval at 95% did not include zero were included. Electrodes are reported in a regular grid array that approximate their position on the scalp.

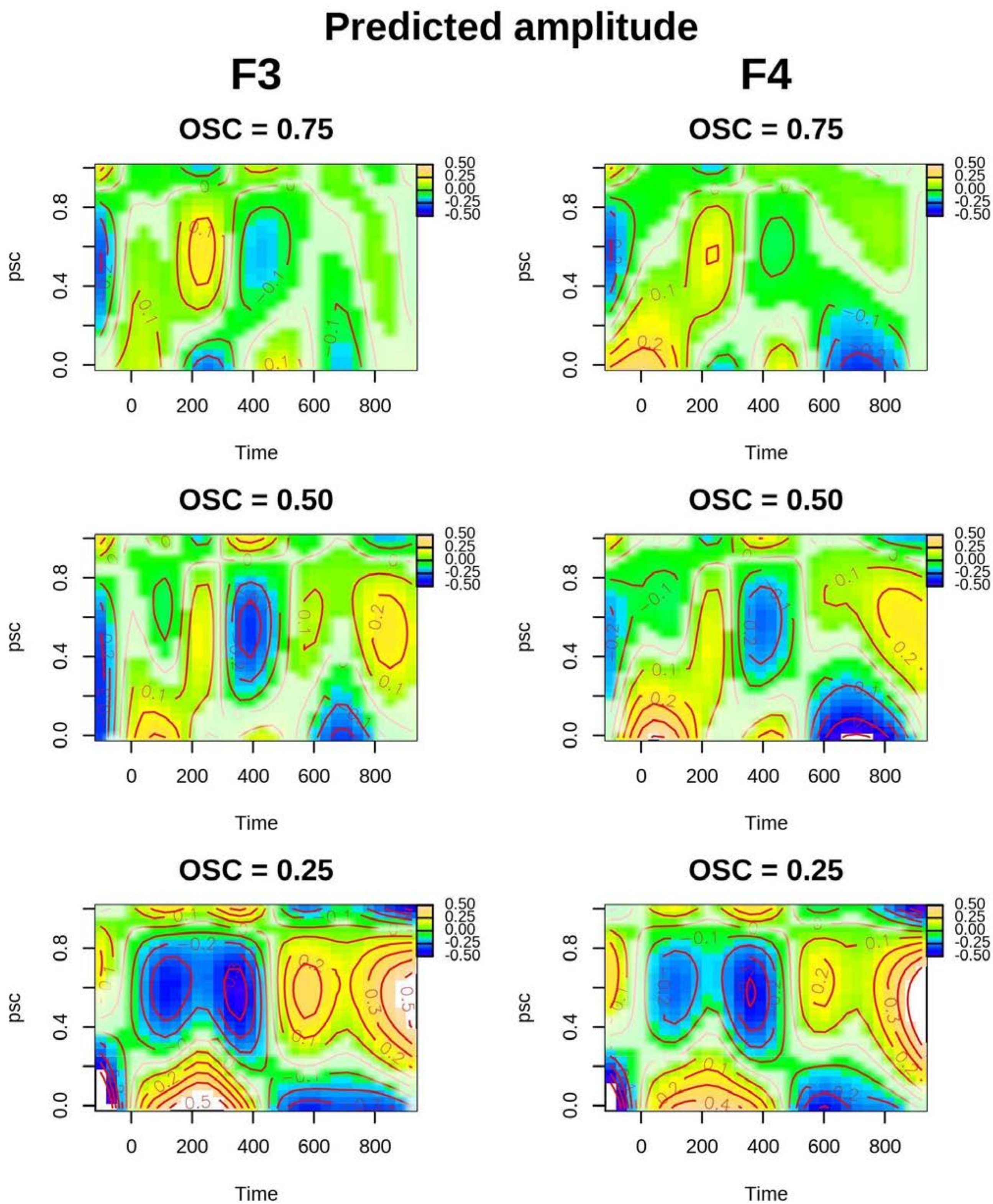


Figure 5. Effect of PSC and OSC over time in frontal electrodes

The figure shows the partial effect of PSC conditioned for three values of OSC (upper panel: OSC set to 0.75, middle panel: OSC set to 0.5, lower panel: OSC set to 0.25), in two frontal electrodes, F3 and F4. The contour map is a topographic representation of the partial effect with the *itsadug* package default palette: colors toward orange indicate positive values, while colors toward blue indicate negative values, while colors toward green indicate in-between values. Topographic maps have been masked so that only estimates whose confidence interval at 95% did not include zero were included. Effect on all timepoints can be inspected with this app https://giorgioarcara.shinyapps.io/ERP-OSC-PSC

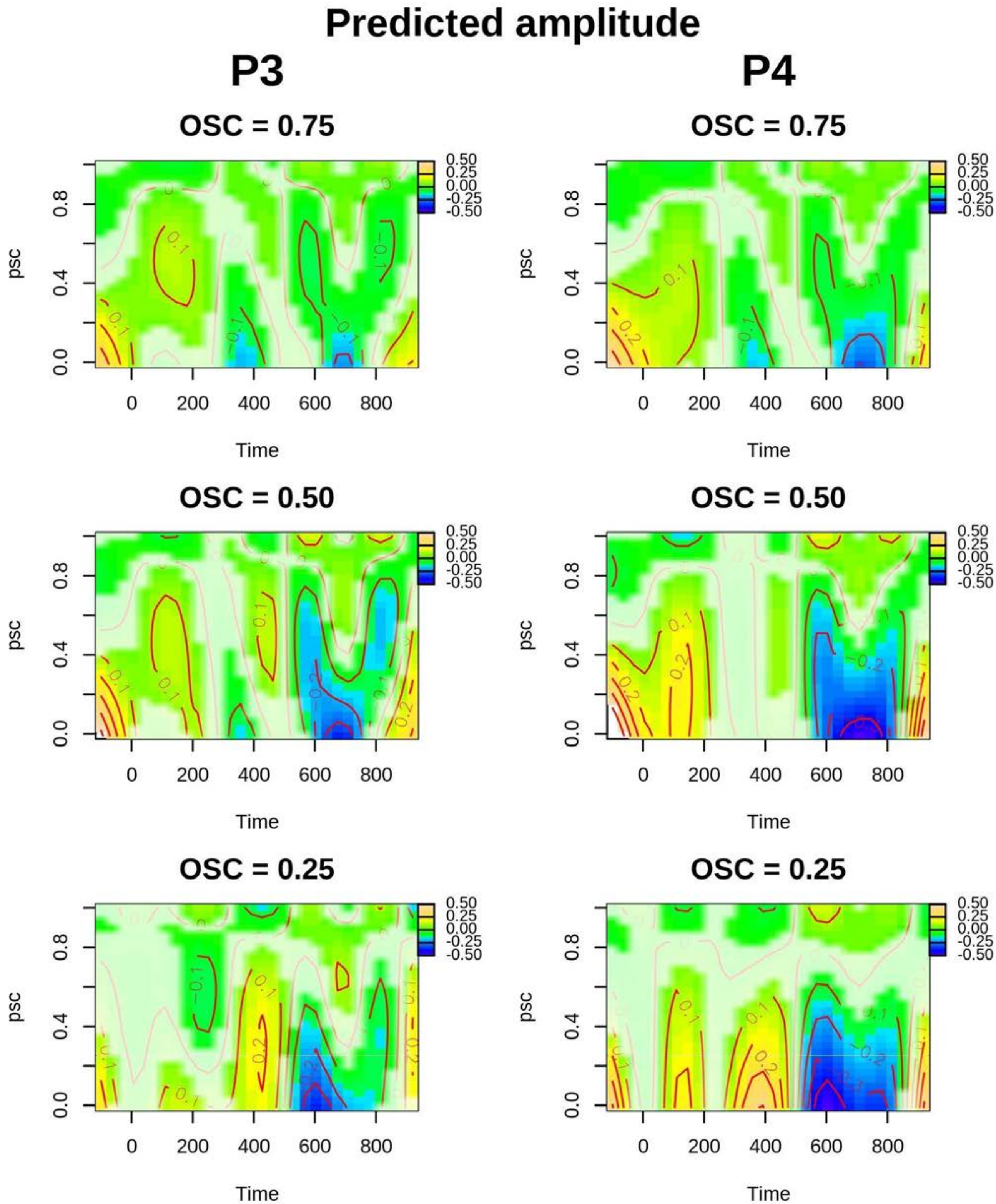


Figure 6. Effect of PSC and OSC over time in parietal electrodes

The figure shows the partial effect of PSC conditioned for three values of OSC (upper panel: OSC set to 0.75, middle panel: OSC set to 0.5, lower panel: OSC set to 0.25), in two parietal electrodes, P3 and P4. The contour map is a topographic representation of the partial effect with the *itsadug* package default palette: colors toward orange indicate positive values, while colors toward blue indicate negative values, while colors toward green indicate in-between values. Topographic maps have been masked so that only estimates whose confidence interval at 95% did not include zero were included. Effect on all timepoints can be inspected with this app https://giorgioarcara.shinyapps.io/ERP-OSC-PSC

4. Discussion

In this study, we used an advanced analytic method (i.e., Generalized additive models, GAM) to evaluate the impact of form-meaning relations – as indexed by OSC and PSC – on participants' electrophysiological response in a lexical decision task with visually presented words. Our results show very early effects of OSC and PSC, already visible at ~100 ms after word presentation. At this time point, the interaction between the two variables showed that, on frontal and fronto-central electrodes, at low values of OSC, higher values of PSC were associated with a larger negativity. Since OSC and PSC capture the relation between form and meaning, their joint effect indicates an early activation of both orthography and phonology, which are immediately mapped into semantics to constrain the recognition process. This early effect is long lasting, being visible up to ~250 ms after word presentation, and is followed by a second effect showing up between ~300 and ~400 ms after word presentation: OSC and PSC interacted at frontal and posterior sites, showing that, on frontal sites, at low values of OSC, middle-to-higher values of PSC were associated to a larger negativity, whereas at posterior sites, at low values of OSC, low values of PSC were associated to a larger positivity.

A bimodal time dynamics for semantic processing characterized by a first very early access plus a second later processing has been also reported by Hauk, Davis, et al. (2006). In their lexical decision study with EEG, the authors found semantic effects picking at 160 ms and then at 314 after word presentation. Also, Dell'Acqua et al. (2010) studied semantic processing by means of the picture-word interference paradigm and ERPs. The authors found semantic effects at two different latencies, i.e., 106 ms and 320 ms, with semantic processing being again characterized by a comparable bimodal time dynamics. These findings show converging evidence that semantic information – as accessed by word forms – contributes to word recognition since the early stages and at multiple times during the process. Suggestively, a similar dual access to semantic information has been proposed for object recognition. According to Clarke and Tyler (2015), semantic information can be accessed within 150 ms from the object presentation, with semantic effects being occasionally visible even earlier than 100 ms (e.g., Clarke et al., 2013). This fast access allows a coarse semantic analysis of the stimulus which is based on semantic information the object shares with many other entities (e.g., *has legs*). The coarse analysis is sufficient for a coarse categorization – e.g., to distinguish a living from a non-living entity. Later on, a more fine-grained analysis allows to access to more specific semantic information, permitting, e.g., to distinguish members belonging to the same category. The visual word recognition system might exploit the same dual mechanism for semantic access: Moving from this perspective, the early and late semantic effect we reported might be associated with a coarse and fine grained semantic analysis, respectively. During the coarse analysis the system might capitalize on the systematic relations between the word form and its meaning to get a first rough hint of the lexical nature of the stimulus. As a matter of fact, the activation of semantic information, although coarse, would be a strong evidence to categorize a printed stimulus as a word and thus accomplish the lexical decision task (for the possibility of a semantic-based lexical decision in a connectionist perspective, see, e.g., Chuang et al., 2020; Harm & Seidenberg, 2004; Plaut, 1997). The late effect, instead, would reflect a more fine-grained semantic analysis which is clearly detectable in our results at 400 ms. This time window is fully in line with evidence from the extensive ERP literature on visual word recognition, in which semantic effects are consistently reported at ~400 ms and typically interpreted as N400 modulations, a component indexing semantic processing (for reviews, see, e.g., Grainger & Holcomb, 2009; Kutas & Federmeier, 2011; Lau et al., 2008; in Dufau et al. (2015) the effect of concreteness emerged ~300 ms; in Sassenhagen and Fiebach semantic properties encoded by distributional word vectors affect ERP responses in a 300-500ms time window). In a recent work using a connectionist model to explore the mechanism underlying the N400, Rabovsky and McRae (2014) suggested that "N400 amplitudes seem to depend crucially on the similarity between actual observations and implicit anticipations based on represented occurrence probabilities as extracted from previously experienced regularities" (p. 83). In other words, the N400 amplitude would depend on the implicit prediction error, which is the mismatch between the external world and its internal model continuously updated by the brain. In such a perspective, the reliability of information at the form (e.g., orthographic) level may affect prediction error at semantic level: when the link between form and meaning is weak (i.e., when OSC is low), implicit expectations of irrelevant information is high as orthographic information does not point toward any well-defined portion of the semantic space.

It is worth nothing that the time course of the early interaction between OSC and PSC is typically associated to visuo-orthographic processing, as indexed by effects of letter length, bigram frequency, and word frequency, all starting within the first 200 ms after stimulus presentation (e.g., Hauk, Davis et al., 2006; 2009; Laszlo & Federmeier, 2014). Dufau et al. (2015) – who analyzed the same dataset we used here – reported that the effect of letter length arose at ~100 ms after stimulus presentation, and was immediately followed by that of word frequency. It must be noted that the interpretation of the effect of frequency is currently debated. Indeed, frequency might impact word recognition not because it reflects experience with the presented word form (as implicitly assumed in Dufau et al.), but rather as an epiphenomenon of conceptual familiarity (Baayen, Feldman and

Schreuder, 2006). The nature of the frequency effect will hence be intrinsically semantic, capturing the ease of accessibility to the word-denoted concept. In such a perspective, the findings by Dufau et al. (2015) may be (at least in part) reconciled with our results: their reported frequency effect might reflect the same conceptual-access process that emerges, in our analyses, as an early interaction between OSC and PSC.

From a modeling perspective, the presence of an interaction between OSC and PSC, as well as its time dynamics, clearly supports the view that word meaning is accessed by means of two distinct and interactive paths – i.e., the orthography-to-meaning and the orthography-to-phonology-to-meaning path – which mutually contribute to word recognition. The time dynamics of the effect indicates that, as soon as a printed string is presented, both paths become immediately active: Within 100 ms, the reader starts accessing orthographic, phonological and semantic information, which can be all used to reach a decision. Of course, this does not mean that in this short time frame a complete speech-level phonetic process takes place (i.e., involving the full reactivation of the auditory form of a word) or that the visual information related to the printed word is reactivated; rather coarse-grained, abstract phonological and orthographic representations become available and inform word processing (for very early effects (i.e., within 100 ms) of phonological activation in printed word processing, see, e.g., Klein et al., 2015; Wheat, Cornelissen, Frost, & Hanse, 2010). At the cognitive level, our findings fit with Harm and Seidenberg's (2004) connectionist model of word recognition, in which activations of both orthography and phonology directly co-occur for accessing word meaning. Of particular interest, in this model, meaning is determined by both paths simultaneously, with semantic patterns reflecting "the joint effects of input from different sources" (p. 663). Thus, as soon as orthographic information becomes available, it directly spreads to semantics and phonology, and phonology, in turn, spreads to semantics itself. As a result, orthography, phonology, and semantics all contribute in parallel to recognize a printed stimulus as a word since the early stage of processing. A further promising interpretative framework comes from the recent work by Chuang et al. (2020), in which the mapping between form and meaning is linearly modeled via regression. Chuang's simulations show that the consistency between the semantic vectors estimated by the two possible reading routes (orthography-based vs. phonology-based) determines lexical processing. This result can be interpreted as evidence for the two routes interacting during stimulus evaluation, rather than learning.

There are some technical details that are important to note for a correct interpretation of our results. The first is that the scalp distribution of effects can be dependent on the reference used. In this case, we kept the left-mastoid reference (as from the recording), but a different reference could lead to different spatial distribution of the effects. This is not a specific issue of GAMs, but of any EEG experiment (Luck, 2014). The second one is that each effect that is discussed should not be considered alone, but always in the context of the overall model. In particular, this means that the evidenced effects of OSC and PSC over time were found in the context of a model that also accounts for the impact of number of letters and frequency. This latter detail could explain small discrepancies with other studies on the Kiloword dataset, which differently modeled the effects of predictors and took into account different variables (e.g., see results obtained with the same dataset in Dufau et al., 2015). Finally, it is important to stress that the current implementation of GAM used in the study does not allow to identify when a particular interaction becomes significant (or not): this would require the development of a GAM implementation, which is beyond the scope of the present study. Rather, the model we used here helps at capturing the time-related dynamics of complex interactions (such as the one between OSC, PSC, and Time) suggesting when higher and more relevant effects were evidenced, and can provide indications that may guide future research (and analysis) focusing on directly testing when effects arise.

To conclude, in the present study we investigated the time-course of semantic processing in visual word recognition by using neurally-plausible distributional semantic measures and state-of-the art data-analytic techniques, and capitalizing on lexical decision data from an ERP mega-study. We reported a bimodal timed dynamic of semantic processing with a very early interaction effect of OSC and PSC – likely associated with a first coarse semantic analysis – and a later interaction effect – likely associated with a fine-grained semantic access. These findings indicate that the recognition system is highly sensitive to form-meaning relations established at different levels of granularity, confirming the central role of systematicity (i.e., the statistical relationship between the patterns of letter/sound for a group of words and their usage, Dingemanse et al., 2015) in supporting word identification. Remarkably, investigating such a process was made possible by measures rooted in distributional semantics. Distributional models offer a convenient and increasingly popular method to quantitatively characterize semantic memory, which builds on cognitively-plausible association mechanisms and was shown to provide meaningful predictions for a number of psychological phenomena (Gunther, Rinaldi & Marelli, 2019). The learning procedures leading to semantic representations in distributional models might be the very same that lead to the form-meaning associations captured by the PSC-by-OSC interaction, and explain the remarkable readers' ability to exploit systematicity (Baayen et al., 2011; Chuang et al., 2019), i.e., the stable link between word's form properties and meaning based on statistical co-occurrence (Dingemanse et al., 2015). Systematicity may be helpful for readers: orthographic (and phonological) similarity among words with similar meaning may support both word learning and lexical organization, offering a clustering principle to group words.

Funding

GA and SL were funded by the Italian Ministry of Health.

References

Amenta, S., Crepaldi, D., & Marelli, M. (2020). Author accepted manuscript: Consistency measures individuate dissociating semantic modulations in priming paradigms: A new look on semantics in the processing of (complex) words. Quarterly Journal of Experimental Psychology, 1747021820927663.

Amenta, S., Marelli, M., & Sulpizio, S. (2017). From sound to meaning: Phonology-to-Semantics mapping in visual word recognition. *Psychonomic bulletin & review*, *24*, 887-893.

Amsel, B. D., Urbach, T. P., & Kutas, M. (2013). Alive and grasping: Stable and rapid semantic access to an object category but not object graspability. *NeuroImage*, *77*, 1-13.

Arcara, G., & Petrova, A. (2017). erpR: event-Related Potentials (ERP) Analysis, Graphics and Utility Functions.

Assadollahi, R., & Pulvermüller, F. (2003). Early influences of word length and frequency: a group study using MEG. *Neuroreport*, *14*, 1183-1187.

Baroni,M., Dinu, G., & Kruszewski, G. (2014b). Don't count, predict! A systematic comparison of context-counting vs. context-predicting semantic vectors. In Proceedings of the Association for Computational Linguistics (ACL) (pp. 238–247). New York, NY: ACM Press.

Baayen, R. H., Kuperman, V., & Bertram, R. (2010). Frequency effects in compound processing. In Cross-Disciplinary Issues in Compounding (pp. 257-270). John Benjamins.

Baayen, R. H., Milin, P., Đurđević, D. F., Hendrix, P., & Marelli, M. (2011). An amorphous model for morphological processing in visual comprehension based on naive discriminative learning. *Psychological review*, *118*, 438-481.

Baayen, R. H., Piepenbrock, R., & Gulikers, L. (1995). *The CELEX lexical database* (CD Rom). University of Pennsylvania, Philadelphia, PA: Linguistic Data Consortium.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the royal statistical society. Series B (Methodological)*, 289-300.

Bentin, S., Mouchetant-Rostaing, Y., Giard, M. H., Echallier, J. F., & Pernier, J. (1999). ERP manifestations of processing printed words at different psycholinguistic levels: time course and scalp distribution. *Journal of Cognitive Neuroscience*, *11*, 235-260.

Carrasco-Ortiz, H., Midgley, K. J., Grainger, J., & Holcomb, P. J. (2017). Interactions in the neighborhood: Effects of orthographic and phonological neighbors on N400 amplitude. *Journal of Neurolinguistics*, *41*, 1-10.

Chan, A. M., Baker, J. M., Eskandar, E., Schomer, D., Ulbert, I., Marinkovic, K., ... & Halgren, E. (2011). First-pass selectivity for semantic categories in human anteroventral temporal lobe. *Journal of Neuroscience*, *31*, 18119-18129.

Chen, Y., Davis, M. H., Pulvermüller, F., & Hauk, O. (2015). Early visual word processing is flexible: Evidence from spatiotemporal brain dynamics. *Journal of Cognitive Neuroscience*, *27*, 1738-1751.

Chuang, Y. Y., Vollmer, M. L., Shafaei-Bajestan, E., Gahl, S., Hendrix, P., & Baayen, R. H. (2020). The processing of pseudoword form and meaning in production and comprehension: A computational modeling approach using Linear Discriminative Learning. *Behavior Research Methods.* Advanced online publication. https://doi.org/10.3758/s13428-020-01356-w

Clarke, A., & Tyler, L. K. (2015). Understanding what we see: how we derive meaning from vision. *Trends in cognitive sciences*, *19*, 677-687.

Clarke, A., Taylor, K. I., Devereux, B., Randall, B., & Tyler, L. K. (2013). From perception to conception: how meaningful objects are processed over time. *Cerebral Cortex*, *23*, 187-197.

De Cat, C., Klepousniotou, E., and Baayen, R. H. (2015). Representational deficit or processing effect?

An electrophysiological study of noun-noun compound processing by very advanced L2 speakers of English. *Frontiers in Psychology, 6,* 77.

Dehaene, S., & Cohen, L. (2011). The unique role of the visual word form area in reading. *Trends in cognitive sciences*, *15*, 254-262.

Dingemanse, M., Blasi, D. E., Lupyan, G., Christiansen, M. H., & Monaghan, P. (2015). Arbitrariness, iconicity, and systematicity in language. *Trends in cognitive sciences*, *19*, 603-615.

Dufau, S., Grainger, J., Midgley, K. J., & Holcomb, P. J. (2015). A thousand words are worth a picture: Snapshots of printed-word processing in an event-related potential megastudy. *Psychological science, 26,* 1887-1897.

Dell'Acqua, R., Sessa, P., Peressotti, F., Mulatti, C., Navarrete, E., & Grainger, J. (2010). ERP evidence for ultra-fast semantic processing in the picture–word interference paradigm. *Frontiers in psychology*, *1*, 177.

Grainger, J., & Holcomb, P. J. (2009). Watching the word go by: On the time-course of component processes in visual word recognition. *Language and linguistics compass*, *3*, 128-156.

Grainger, J., & Ziegler, J. (2011). A dual-route approach to orthographic processing. *Frontiers in psychology*, *2*, 54.

Groppe, D. M., Urbach, T. P., & Kutas, M. (2011). Mass univariate analysis of event-related brain potentials/fields I: A critical tutorial review. *Psychophysiology*, *48*, 1711-1725.

Günther, F., Rinaldi, L., & Marelli, M. (2019). Vector-space models of semantic representation from a cognitive perspective: A discussion of common misconceptions. *Perspectives on Psychological Science*, *14*(6), 1006-1033.

Harrell Jr, F. E. (2015). *Regression modeling strategies: with applications to linear models, logistic and ordinal regression, and survival analysis*. Springer-Verlag: New York, NY.

Harm, M. W., & Seidenberg, M. S. (2004). Computing the meanings of words in reading: Cooperative division of labor between visual and phonological processes. Psychological Review, 111, 662–720.

Hauk, O., Coutout, C., Holden, A., & Chen, Y. (2012). The time-course of single-word reading: evidence from fast behavioral and brain responses. *NeuroImage*, *60*, 1462-1477.

Hauk, O., Davis, M. H., Ford, M., Pulvermüller, F., & Marslen-Wilson, W. D. (2006). The time course of visual word recognition as revealed by linear regression analysis of ERP data. *Neuroimage*, *30*, 1383-1400.

Hauk, O., Patterson, K., Woollams, A., Watling, L., Pulvermüller, F., & Rogers, T. T. (2006). [Q:] When would you prefer a SOSSAGE to a SAUSAGE?: At about 100 msec. ERP correlates of orthographic typicality and lexicality in written word recognition. *Journal of Cognitive Neuroscience*, *18*, 818-832.

Hauk, O., Pulvermüller, F., Ford, M., Marslen-Wilson, W. D., & Davis, M. H. (2009). Can I have a quick word? Early electrophysiological manifestations of psycholinguistic processes revealed by event-related regression analysis of the EEG. *Biological psychology*, *80*, 64-74.

Hendrix, P., Bolger, P., and Baayen, R. H. (2016). Distinct ERP signatures of word frequency, phrase frequency, and prototypicality in speech production. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 43,* 28-149.

Jones, M. N., Willits, J., Dennis, S., & Jones, M. (2015). Models of semantic memory. *Oxford handbook of mathematical and computational psychology*, 232-254.

Just, M. A., Cherkassky, V. L., Aryal, S., & Mitchell, T. M. (2010). A neurosemantic theory of concrete noun representation based on the underlying brain codes. *PloS one*, *5*, e8622.

Klein, M., Grainger, J., Wheat, K. L., Millman, R. E., Simpson, M. I., Hansen, P. C., & Cornelissen, P. L. (2015). Early activity in Broca's area during reading reflects fast access to articulatory codes from print. *Cerebral Cortex, 25*, 1715-1723.

Kryuchkova, T., Tucker, B. V., Wurm, L., and Baayen R. H. (2012). Danger and usefulness in auditory lexical processing: evidence from electroencephalography. *Brain and Language 122*, 81-91.

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*, 621-647.

Landauer, T., & Dumais, S. (1997). A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction and repre- sentation of knowledge. *Psychological Review, 104,* 211–240.

Lau, E. F., Phillips, C., & Poeppel, D. (2008). A cortical network for semantics:(de) constructing the N400. *Nature Reviews Neuroscience*, *9*, 920-933.

Lund, K., & Burgess, C. (1996). Producing high-dimensional semantic spaces from lexical cooccurrence. *Behavioural Research Methods, 28,* 203-208.

Laszlo, S., & Federmeier, K. D. (2014). Never seem to find the time: evaluating the physiological time course of visual word recognition with regression analysis of single-item event-related potentials. *Language, Cognition and Neuroscience*, *29*, 642-661.

Louwerse, M., & Hutchinson, S. (2012). Neurological evidence linguistic processes precede perceptual simulation in conceptual processing. *Frontiers in psychology*, *3*, 385.

Luck, S. J. (2014). *An introduction to the event-related potential technique*. MIT press.

Marelli, M., Amenta, S., & Crepaldi, D. (2015). Semantic transparency in free stems: The effect of Orthography-Semantics Consistency in word recognition. *Quarterly Journal of Experimental Psychology, 68,* 1571–1583.

Maurer, U., Schulz, E., Brem, S., van der Mark, S., Bucher, K., Martin, E., & Brandeis, D. (2011). The development of print tuning in children with dyslexia: Evidence from longitudinal ERP data supported by fMRI. *Neuroimage*, *57*, 714-722.

McClelland, J. L. (1979). On the time relations of mental processes: An examination of systems of processes in cascade. *Psychological Review, 86,* 287-330.

Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., & Dean, J. (2013). Distributed representations of words and phrases and their compositionality. In C. J. C. Burges, L. Bottou, M. Welling, Z. Ghahramani, & K. Q. Weinberger (Eds.), Proceedings of the 26th International Conference on Neural Information Processing Systems (Vol. 2, pp. 3111–3119). New York, NY: Curran Associates.

Mitchell, T. M., Shinkareva, S. V., Carlson, A., Chang, K. M., Malave, V. L., Mason, R. A., & Just, M. A. (2008). Predicting human brain activity associated with the meanings of nouns. *Science*, *320*, 1191-1195.

Pereira, F., Lou, B., Pritchett, B., Ritter, S., Gershman, S. J., Kanwisher, N., ... & Fedorenko, E. (2018). Toward a universal decoder of linguistic meaning from brain activation. *Nature communications*, *9*, 963.

Perrone-Bertolotti, M., Kauffmann, L., Pichat, C., Vidal, J. R., & Baciu, M. (2017). Effective connectivity between ventral occipito-temporal and ventral inferior frontal cortex during lexico-semantic processing. A dynamic causal modeling study. *Frontiers in human neuroscience*, *11*, 325.

Plaut, D. C. (1997). Structure and function in the lexical system: Insights from distributed models of word reading and lexical decision. *Language and cognitive processes*, *12*, 765-806.

Price, C. J., & Devlin, J. T. (2011). The interactive account of ventral occipitotemporal contributions to reading. *Trends in cognitive sciences*, *15*, 246-253.

Pulvermüller, F., Shtyrov, Y., & Hauk, O. (2009). Understanding in an instant: neurophysiological evidence for mechanistic language circuits in the brain. *Brain and language*, *110*, 81-94.

Rabovsky, M., & McRae, K. (2014). Simulating the N400 ERP component as semantic network error:

Insights from a feature-based connectionist attractor model of word meaning. *Cognition*, *132*(1), 68–89. https://doi.org/10.1016/j.cognition.2014.03.010

Rabovsky, M., Sommer, W., & Abdel Rahman, R. (2012). The time course of semantic richness effects in visual word recognition. *Frontiers in Human Neuroscience*, *6*, 11.

Sassenhagen, J., & Fiebach, C. J. (2020). Traces of meaning itself: Encoding distributional word vectors in brain activity. *Neurobiology of Language, 1*(1), 54-76.

Schurz, M., Kronbichler, M., Crone, J., Richlan, F., Klackl, J., & Wimmer, H. (2014). Top-down and bottom-up influences on the left ventral occipito-temporal cortex during visual word recognition: An analysis of effective connectivity. *Human brain mapping*, *35*, 1668-1680.

Shallice, T., Warrington, E. K., & McCarthy, R. (1983). Reading without semantics. *The quarterly journal of experimental psychology Section A*, *35*, 111-138.

Siegelman, N., Rueckl, J. G., Lo, J. C. M., Kearns, D. M., Morris, R. D., & Compton, D. L. (2022). Quantifying the regularities between orthography and semantics and their impact on group- and individual-level behavior. *Journal of Experimental Psychology: Learning, Memory, and Cognition. Advance online publication.* https://doi.org/10.1037/xlm0001109

Stites, M. C., & Laszlo, S. (2017). Time will tell: A longitudinal investigation of brain-behavior relationships during reading development. *Psychophysiology*, *54*, 798-808.

Tremblay, A., & Newman, A. J. (2015). Modeling nonlinear relationships in ERP data using mixed-effects regression with R examples. *Psychophysiology*, *52*, 124-139.

Turney, P. D., & Pantel, P. (2010). From frequency to meaning: Vector space models of semantics. *Journal of Artificial Intelligence Research, 37*, 141-188.

Vinckier, F., Dehaene, S., Jobert, A., Dubus, J. P., Sigman, M., & Cohen, L. (2007). Hierarchical coding of letter strings in the ventral stream: dissecting the inner organization of the visual word-form system. *Neuron*, *55*, 143-156.

Wheat, K. L., Cornelissen, P. L., Frost, S. J., & Hansen, P. C. (2010). During visual word recognition, phonology is accessed within 100 ms and may be mediated by a speech production code: evidence from magnetoencephalography. *The Journal of Neuroscience, 30*, 5229-5233.

Wood, S., & Wood, M. S. (2015). Package 'mgcv'. *R package version*, *1*, 29.

Appendix

Table A1. The table shows the results of the linear term for all the models (one for each electrode) fit on Kilo-word data. The first column reports the electrode name; the second column report the term name. The third column reports the effect Estimate (i.e., β); The fourth column reports the Standard Error; the fifth and the sixth columns report the t value and the p-value. The fifth column reports the p-values corrected with False Discovery Rate (FDR) method.

Electrode	term	Estimate	SE	t-value	p-value	p-bonf	p-fdr
T8	(Intercept)	-0.8	0.04	-19.9	<0.001	<0.001	<0.001
CP1	(Intercept)	-1.52	0.05	-29.34	<0.001	<0.001	<0.001
CP6	(Intercept)	-0.77	0.04	-17.64	<0.001	<0.001	<0.001
C3	(Intercept)	-1.33	0.05	-27.56	<0.001	<0.001	<0.001
Cz	(Intercept)	-1.86	0.05	-35.03	<0.001	<0.001	<0.001
C4	(Intercept)	-1.43	0.05	-27.21	<0.001	<0.001	<0.001
P7	(Intercept)	-0.05	0.03	-1.47	0.141	1.00	0.141
Pz	(Intercept)	-1.5	0.05	-27.99	<0.001	<0.001	<0.001
P3	(Intercept)	-0.78	0.05	-17.29	<0.001	<0.001	<0.001
T7	(Intercept)	-0.67	0.04	-17.6	<0.001	<0.001	<0.001
Fp2	(Intercept)	-0.81	0.06	-13.05	<0.001	<0.001	<0.001

FC2	(Intercept)	-1.68	0.05	-32.84	<0.001	<0.001	<0.001
Fp1	(Intercept)	-0.92	0.06	-15.78	<0.001	<0.001	<0.001
CP5	(Intercept)	-0.74	0.04	-18.27	<0.001	<0.001	<0.001
O2	(Intercept)	0.76	0.04	19.05	<0.001	<0.001	<0.001
P8	(Intercept)	0.2	0.04	5.63	<0.001	<0.001	<0.001
F3	(Intercept)	-1.5	0.05	-31.03	<0.001	<0.001	<0.001
Fz	(Intercept)	-1.63	0.05	-30.93	<0.001	<0.001	<0.001
FC1	(Intercept)	-1.7	0.05	-33.96	<0.001	<0.001	<0.001
F4	(Intercept)	-1.51	0.05	-31.14	<0.001	<0.001	<0.001
FC6	(Intercept)	-1.25	0.04	-27.77	<0.001	<0.001	<0.001
P4	(Intercept)	-0.73	0.05	-15.58	<0.001	<0.001	<0.001
CP2	(Intercept)	-1.56	0.05	-29.94	<0.001	<0.001	<0.001
F8	(Intercept)	-1.11	0.04	-24.93	<0.001	<0.001	<0.001
F7	(Intercept)	-1.04	0.04	-26.13	<0.001	<0.001	<0.001
Oz	(Intercept)	-0.15	0.04	-4.02	<0.001	0.002	<0.001
FC5	(Intercept)	-1.21	0.04	-29.47	<0.001	<0.001	<0.001
Fpz	(Intercept)	-1.04	0.06	-18.4	<0.001	<0.001	<0.001

O1	(Intercept)	0.62	0.04	14.92	<0.001	<0.001	<0.001

Table A2. The table shows the results of the smooth terms for all the models (one for each electrode) fit on Kilo-word data. The first column reports the electrode name; the second column reports the smooth term name; The third column reports the Effective degrees of freedom (which is a value related to the number of parameters used to estimate the smooth function), The fourth column reports the Reference degrees of freedom (used to calculated statistics, and p-values); the fifth and the sixth columns report the F value and the p-value. The seventh column reports the p-values corrected with False Discovery Rate (FDR) method.

Electrode	**term**	**Estimated df**	**Reference df**	**F value**	**p-value**	**p-bonf**	**p-fdr**
T8	s(WordFrequency)	1.01	1.01	0.09	0.771	1,00	0.857
	s(Time)	9,00	9,00	9011.79	<0.001	<0.001	<0.001
	s(NumberOfLetters)	1.01	1.01	2.79	0.096	1,00	0.148
	s(psc)	1,00	1,00	0.14	0.713	1,00	0.825
	s(osc)	1,00	1,00	0.51	0.474	1,00	0.603
	ti(WordFrequency,Time)	17.03	17.88	62.44	<0.001	<0.001	<0.001
	ti(NumberOfLetters,Time)	17.54	17.97	78.07	<0.001	<0.001	<0.001
	ti(psc,Time)	15.15	15.84	23.78	<0.001	<0.001	<0.001

	ti(osc,Time)	14.69	15.68	21.19	<0.001	<0.001	<0.001
	ti(psc,osc)	1.77	1.77	0.8	0.428	1,00	0.557
	ti(psc,osc,Time)	31.26	34.27	17.52	<0.001	<0.001	<0.001
	s(WORD)	677.61	683,00	147.16	<0.001	<0.001	<0.001
CP1	s(WordFrequency)	1,00	1,00	0.14	0.713	1,00	0.825
	s(Time)	9,00	9,00	17435.91	<0.001	<0.001	<0.001
	s(NumberOfLetters)	1.01	1.01	2.2	0.139	1,00	0.21
	s(psc)	1,00	1,00	0.12	0.733	1,00	0.83
	s(osc)	1,00	1,00	0,00	0.994	1,00	0.996
	ti(WordFrequency,Time)	17.02	17.88	51.27	<0.001	<0.001	<0.001
	ti(NumberOfLetters,Time)	17.26	17.93	126.43	<0.001	<0.001	<0.001
	ti(psc,Time)	15.31	15.91	26.08	<0.001	<0.001	<0.001
	ti(osc,Time)	15.28	15.88	33.86	<0.001	<0.001	<0.001
	ti(psc,osc)	1.02	1.02	0,00	0.995	1,00	0.996
	ti(psc,osc,Time)	30.16	33.71	7.66	<0.001	<0.001	<0.001
	s(WORD)	678.86	683,00	166.34	<0.001	<0.001	<0.001
CP6	s(WordFrequency)	1.01	1.01	0.01	0.903	1,00	0.941

	s(Time)	9,00	9,00	5923.03	<0.001	<0.001	<0.001
	s(NumberOfLetters)	1.01	1.01	2.74	0.099	1,00	0.152
	s(psc)	1,00	1,00	0.01	0.918	1,00	0.95
	s(osc)	1.01	1.01	0.21	0.641	1,00	0.772
	ti(WordFrequency,Time)	16.96	17.86	59.23	<0.001	<0.001	<0.001
	ti(NumberOfLetters,Time)	17.48	17.97	123.91	<0.001	<0.001	<0.001
	ti(psc,Time)	15.31	15.91	24.86	<0.001	<0.001	<0.001
	ti(osc,Time)	14.69	15.59	27.46	<0.001	<0.001	<0.001
	ti(psc,osc)	1.1	1.1	0.09	0.824	1,00	0.887
	ti(psc,osc,Time)	31.54	34.74	13.27	<0.001	<0.001	<0.001
	s(WORD)	678.17	683,00	144.63	<0.001	<0.001	<0.001
C3	s(WordFrequency)	1,00	1,00	0.37	0.542	1,00	0.673
	s(Time)	9,00	9,00	19555.86	<0.001	<0.001	<0.001
	s(NumberOfLetters)	1,00	1,00	1.31	0.252	1,00	0.36
	s(psc)	1,00	1,00	0.2	0.653	1,00	0.775
	s(osc)	1,00	1,00	0,00	0.948	1,00	0.976
	ti(WordFrequency,Time)	17.39	17.95	50.54	<0.001	<0.001	<0.001

	ti(NumberOfLetters,Time)	17.2	17.92	109.18	<0.001	<0.001	<0.001
	ti(psc,Time)	15.07	15.83	20.5	<0.001	<0.001	<0.001
	ti(osc,Time)	15.21	15.84	51.2	<0.001	<0.001	<0.001
	ti(psc,osc)	1.02	1.02	0.04	0.843	1,00	0.902
	ti(psc,osc,Time)	30.44	33.85	12.42	<0.001	<0.001	<0.001
	s(WORD)	678.9	683,00	167.99	<0.001	<0.001	<0.001
Cz	s(WordFrequency)	1.01	1.01	0.08	0.782	1,00	0.859
	s(Time)	9,00	9,00	23831.26	<0.001	<0.001	<0.001
	s(NumberOfLetters)	1.01	1.01	2.77	0.096	1,00	0.148
	s(psc)	1,00	1,00	0.14	0.711	1,00	0.825
	s(osc)	1.01	1.01	0.13	0.724	1,00	0.828
	ti(WordFrequency,Time)	17.37	17.95	59.63	<0.001	<0.001	<0.001
	ti(NumberOfLetters,Time)	17.02	17.88	125.46	<0.001	<0.001	<0.001
	ti(psc,Time)	15.22	15.89	19.63	<0.001	<0.001	<0.001
	ti(osc,Time)	15.26	15.86	38.39	<0.001	<0.001	<0.001
	ti(psc,osc)	1.09	1.09	0.52	0.477	1,00	0.603
	ti(psc,osc,Time)	30.97	34.12	12.63	<0.001	<0.001	<0.001

	s(WORD)	677.98	683,00	139.02	<0.001	<0.001	<0.001
C4	s(WordFrequency)	1,00	1,00	0.05	0.816	1,00	0.882
	s(Time)	9,00	9,00	16377.35	<0.001	<0.001	<0.001
	s(NumberOfLetters)	1,00	1,00	2.89	0.089	1,00	0.139
	s(psc)	1,00	1,00	0.09	0.761	1,00	0.851
	s(osc)	1,00	1,00	0.28	0.6	1,00	0.725
	ti(WordFrequency,Time)	17.2	17.92	54.99	<0.001	<0.001	<0.001
	ti(NumberOfLetters,Time)	17.08	17.89	112.57	<0.001	<0.001	<0.001
	ti(psc,Time)	15.05	15.81	22.41	<0.001	<0.001	<0.001
	ti(osc,Time)	14.23	15.41	24.75	<0.001	<0.001	<0.001
	ti(psc,osc)	1.02	1.02	0.12	0.723	1,00	0.828
	ti(psc,osc,Time)	29.93	33.37	10.88	<0.001	<0.001	<0.001
	s(WORD)	678.82	683,00	164.73	<0.001	<0.001	<0.001
P7	s(WordFrequency)	1,00	1,00	1.14	0.285	1,00	0.399
	s(Time)	9,00	9,00	2108.81	<0.001	<0.001	<0.001
	s(NumberOfLetters)	1.02	1.02	0.14	0.731	1,00	0.83
	s(psc)	1,00	1,00	0.2	0.653	1,00	0.775

	s(osc)	1,00	1,00	0,00	0.976	1,00	0.989
	ti(WordFrequency,Time)	17.12	17.9	147.2	<0.001	<0.001	<0.001
	ti(NumberOfLetters,Time)	17.44	17.96	63.54	<0.001	<0.001	<0.001
	ti(psc,Time)	15.55	15.96	30.83	<0.001	<0.001	<0.001
	ti(osc,Time)	15.51	15.94	33.77	<0.001	<0.001	<0.001
	ti(psc,osc)	1.02	1.02	0.33	0.566	1,00	0.697
	ti(psc,osc,Time)	31.67	34.88	14.73	<0.001	<0.001	<0.001
	s(WORD)	678.61	683,00	156.83	<0.001	<0.001	<0.001
Pz	s(WordFrequency)	1,00	1,00	1.05	0.304	1,00	0.419
	s(Time)	9,00	9,00	11425.04	<0.001	<0.001	<0.001
	s(NumberOfLetters)	1,00	1,00	4.49	0.034	1,00	0.056
	s(psc)	1,00	1,00	0.1	0.759	1,00	0.851
	s(osc)	1,00	1,00	0.07	0.793	1,00	0.866
	ti(WordFrequency,Time)	17.26	17.93	70.88	<0.001	<0.001	<0.001
	ti(NumberOfLetters,Time)	16.78	17.81	144.9	<0.001	<0.001	<0.001
	ti(psc,Time)	15.07	15.79	26.39	<0.001	<0.001	<0.001
	ti(osc,Time)	15.26	15.87	32.74	<0.001	<0.001	<0.001

	ti(psc,osc)	1.04	1.04	0,00	0.978	1,00	0.989
	ti(psc,osc,Time)	30.67	34.08	12.32	<0.001	<0.001	<0.001
	s(WORD)	678.34	683,00	147.75	<0.001	<0.001	<0.001
P3	s(WordFrequency)	1.01	1.01	0,00	0.966	1,00	0.983
	s(Time)	9,00	9,00	7175.85	<0.001	<0.001	<0.001
	s(NumberOfLetters)	1.01	1.01	1.07	0.303	1,00	0.419
	s(psc)	1.01	1.01	0.02	0.883	1,00	0.931
	s(osc)	1.01	1.01	0.07	0.797	1,00	0.866
	ti(WordFrequency,Time)	17.11	17.9	70.67	<0.001	<0.001	<0.001
	ti(NumberOfLetters,Time)	17.28	17.94	114.74	<0.001	<0.001	<0.001
	ti(psc,Time)	15.48	15.94	28.98	<0.001	<0.001	<0.001
	ti(osc,Time)	15.42	15.93	35.7	<0.001	<0.001	<0.001
	ti(psc,osc)	1.05	1.05	0.39	0.534	1,00	0.666
	ti(psc,osc,Time)	31.35	34.64	10.66	<0.001	<0.001	<0.001
	s(WORD)	678.57	683,00	156.49	<0.001	<0.001	<0.001
T7	s(WordFrequency)	1,00	1,00	3.44	0.064	1,00	0.102
	s(Time)	9,00	9,00	12864.33	<0.001	<0.001	<0.001

	s(NumberOfLetters)	1.01	1.01	0.19	0.679	1,00	0.798
	s(psc)	1.01	1.01	1.28	0.262	1,00	0.371
	s(osc)	1,00	1,00	0.01	0.911	1,00	0.946
	ti(WordFrequency,Time)	17.26	17.93	48.97	<0.001	<0.001	<0.001
	ti(NumberOfLetters,Time)	17.48	17.97	66.57	<0.001	<0.001	<0.001
	ti(psc,Time)	15.44	15.93	47.1	<0.001	<0.001	<0.001
	ti(osc,Time)	15.4	15.91	50.11	<0.001	<0.001	<0.001
	ti(psc,osc)	1.52	1.52	0.2	0.645	1,00	0.774
	ti(psc,osc,Time)	33.93	35.68	31.35	<0.001	<0.001	<0.001
	s(WORD)	678.57	683,00	172.55	<0.001	<0.001	<0.001
Fp2	s(WordFrequency)	1,00	1,00	7.45	0.006	1,00	0.011
	s(Time)	9,00	9,00	15552,00	<0.001	<0.001	<0.001
	s(NumberOfLetters)	1,00	1,00	4.11	0.042	1,00	0.07
	s(psc)	1,00	1,00	1.68	0.195	1,00	0.289
	s(osc)	1,00	1,00	3.51	0.062	1,00	0.099
	ti(WordFrequency,Time)	17.55	17.97	150.97	<0.001	<0.001	<0.001
	ti(NumberOfLetters,Time)	17.19	17.93	55.21	<0.001	<0.001	<0.001

	ti(psc,Time)	14.99	15.83	22.55	<0.001	<0.001	<0.001
	ti(osc,Time)	15.29	15.89	47.41	<0.001	<0.001	<0.001
	ti(psc,osc)	3.25	3.25	1.76	0.189	1,00	0.281
	ti(psc,osc,Time)	31.17	34.03	38.96	<0.001	<0.001	<0.001
	s(WORD)	676.68	683,00	165.59	<0.001	<0.001	<0.001
FC2	s(WordFrequency)	1,00	1,00	0.51	0.474	1,00	0.603
	s(Time)	9,00	9,00	21009.88	<0.001	<0.001	<0.001
	s(NumberOfLetters)	1,00	1,00	2.31	0.128	1,00	0.195
	s(psc)	1,00	1,00	0.98	0.321	1,00	0.435
	s(osc)	1,00	1,00	0.06	0.807	1,00	0.875
	ti(WordFrequency,Time)	17.56	17.97	96.08	<0.001	<0.001	<0.001
	ti(NumberOfLetters,Time)	17.2	17.92	101.03	<0.001	<0.001	<0.001
	ti(psc,Time)	14.89	15.74	16.53	<0.001	<0.001	<0.001
	ti(osc,Time)	14.99	15.77	31.96	<0.001	<0.001	<0.001
	ti(psc,osc)	1.02	1.02	1.09	0.296	1,00	0.411
	ti(psc,osc,Time)	31.24	34.24	16.48	<0.001	<0.001	<0.001
	s(WORD)	678.03	683,00	138.34	<0.001	<0.001	<0.001

Fp1	s(WordFrequency)	1,00	1,00	1.76	0.185	1,00	0.277
	s(Time)	9,00	9,00	19878.27	<0.001	<0.001	<0.001
	s(NumberOfLetters)	1,00	1,00	6.54	0.01	1,00	0.017
	s(psc)	1,00	1,00	0.33	0.566	1,00	0.697
	s(osc)	1,00	1,00	1.32	0.252	1,00	0.36
	ti(WordFrequency,Time)	17.51	17.97	151.91	<0.001	<0.001	<0.001
	ti(NumberOfLetters,Time)	17.37	17.95	74.3	<0.001	<0.001	<0.001
	ti(psc,Time)	14.99	15.82	20.83	<0.001	<0.001	<0.001
	ti(osc,Time)	14.93	15.72	43.45	<0.001	<0.001	<0.001
	ti(psc,osc)	1.43	1.44	0.44	0.688	1,00	0.807
	ti(psc,osc,Time)	32.48	34.97	27.61	<0.001	<0.001	<0.001
	s(WORD)	678.94	683,00	187.32	<0.001	<0.001	<0.001
CP5	s(WordFrequency)	1,00	1,00	0.39	0.534	1,00	0.666
	s(Time)	9,00	9,00	9992.98	<0.001	<0.001	<0.001
	s(NumberOfLetters)	1.01	1.01	1.05	0.312	1,00	0.427
	s(psc)	1,00	1,00	0.02	0.894	1,00	0.934
	s(osc)	1,00	1,00	0,00	0.951	1,00	0.976

	ti(WordFrequency,Time)	16.87	17.84	67.18	<0.001	<0.001	<0.001
	ti(NumberOfLetters,Time)	17.38	17.95	116.95	<0.001	<0.001	<0.001
	ti(psc,Time)	15.51	15.95	30.74	<0.001	<0.001	<0.001
	ti(osc,Time)	15.46	15.93	48.15	<0.001	<0.001	<0.001
	ti(psc,osc)	1.05	1.05	0.02	0.934	1,00	0.964
	ti(psc,osc,Time)	32.14	35.09	9.7	<0.001	<0.001	<0.001
	s(WORD)	679.24	683,00	184.74	<0.001	<0.001	<0.001
O2	s(WordFrequency)	1.01	1.01	0.03	0.864	1,00	0.919
	s(Time)	8.99	9,00	10630.81	<0.001	<0.001	<0.001
	s(NumberOfLetters)	1.87	1.87	7.38	0.007	1,00	0.012
	s(psc)	1.01	1.01	0.08	0.775	1,00	0.858
	s(osc)	1.01	1.01	0.2	0.648	1,00	0.775
	ti(WordFrequency,Time)	16.96	17.87	75.19	<0.001	<0.001	<0.001
	ti(NumberOfLetters,Time)	17.35	17.95	89.98	<0.001	<0.001	<0.001
	ti(psc,Time)	14.8	15.74	14.59	<0.001	<0.001	<0.001
	ti(osc,Time)	15.36	15.9	25.84	<0.001	<0.001	<0.001
	ti(psc,osc)	1.42	1.42	0.73	0.581	1,00	0.709

	ti(psc,osc,Time)	32.36	35.18	12.75	<0.001	<0.001	<0.001
	s(WORD)	672.54	683,00	73.58	<0.001	<0.001	<0.001
P8	s(WordFrequency)	1.01	1.01	1.1	0.294	1,00	0.409
	s(Time)	8.99	9,00	1575.4	<0.001	<0.001	<0.001
	s(NumberOfLetters)	1.91	1.91	9.27	0.002	0.548	0.003
	s(psc)	1,00	1,00	0.62	0.433	1,00	0.56
	s(osc)	1.01	1.01	0.55	0.455	1,00	0.584
	ti(WordFrequency,Time)	16.99	17.87	81.03	<0.001	<0.001	<0.001
	ti(NumberOfLetters,Time)	17.6	17.98	114.14	<0.001	<0.001	<0.001
	ti(psc,Time)	15.26	15.9	22.06	<0.001	<0.001	<0.001
	ti(osc,Time)	14.52	15.53	18.86	<0.001	<0.001	<0.001
	ti(psc,osc)	1.69	1.69	2.03	0.258	1,00	0.367
	ti(psc,osc,Time)	34.09	35.74	21.05	<0.001	<0.001	<0.001
	s(WORD)	675.25	683,00	110.13	<0.001	<0.001	<0.001
F3	s(WordFrequency)	1,00	1,00	0.93	0.335	1,00	0.452
	s(Time)	9,00	9,00	22196.73	<0.001	<0.001	<0.001
	s(NumberOfLetters)	1,00	1,00	2.6	0.106	1,00	0.163

	s(psc)	1,00	1,00	1.83	0.176	1,00	0.265
	s(osc)	1,00	1,00	0.02	0.892	1,00	0.934
	ti(WordFrequency,Time)	17.64	17.98	119.12	<0.001	<0.001	<0.001
	ti(NumberOfLetters,Time)	17.4	17.96	82.36	<0.001	<0.001	<0.001
	ti(psc,Time)	14.48	15.56	13.7	<0.001	<0.001	<0.001
	ti(osc,Time)	15.01	15.79	39.46	<0.001	<0.001	<0.001
	ti(psc,osc)	1.04	1.04	1.16	0.286	1,00	0.399
	ti(psc,osc,Time)	31.93	34.79	22.34	<0.001	<0.001	<0.001
	s(WORD)	678.15	683,00	141.93	<0.001	<0.001	<0.001
Fz	s(WordFrequency)	1,00	1,00	1.33	0.248	1,00	0.358
	s(Time)	9,00	9,00	21108.02	<0.001	<0.001	<0.001
	s(NumberOfLetters)	1,00	1,00	1.66	0.197	1,00	0.29
	s(psc)	1,00	1,00	1.54	0.213	1,00	0.312
	s(osc)	1,00	1,00	1.03	0.313	1,00	0.427
	ti(WordFrequency,Time)	17.61	17.98	133.67	<0.001	<0.001	<0.001
	ti(NumberOfLetters,Time)	17.37	17.95	86.63	<0.001	<0.001	<0.001
	ti(psc,Time)	14.74	15.72	13.41	<0.001	<0.001	<0.001

	ti(osc,Time)	15.25	15.87	40.03	<0.001	<0.001	<0.001
	ti(psc,osc)	1.04	1.04	0.79	0.382	1,00	0.505
	ti(psc,osc,Time)	31.13	34.3	18.15	<0.001	<0.001	<0.001
	s(WORD)	678.21	683,00	143.85	<0.001	<0.001	<0.001
FC1	s(WordFrequency)	1,00	1,00	3.16	0.075	1,00	0.119
	s(Time)	9,00	9,00	22941.21	<0.001	<0.001	<0.001
	s(NumberOfLetters)	1,00	1,00	1.22	0.269	1,00	0.379
	s(psc)	1,00	1,00	0.65	0.421	1,00	0.55
	s(osc)	1,00	1,00	0.08	0.783	1,00	0.859
	ti(WordFrequency,Time)	17.6	17.98	106.35	<0.001	<0.001	<0.001
	ti(NumberOfLetters,Time)	17.35	17.95	100.77	<0.001	<0.001	<0.001
	ti(psc,Time)	14.87	15.77	14.2	<0.001	<0.001	<0.001
	ti(osc,Time)	15.01	15.78	39.09	<0.001	<0.001	<0.001
	ti(psc,osc)	1.04	1.04	0.79	0.384	1,00	0.507
	ti(psc,osc,Time)	31.05	34.23	15.65	<0.001	<0.001	<0.001
	s(WORD)	677.98	683,00	137.12	<0.001	<0.001	<0.001
F4	s(WordFrequency)	1,00	1,00	0.12	0.731	1,00	0.83

	s(Time)	9,00	9,00	20280.88	<0.001	<0.001	<0.001
	s(NumberOfLetters)	1,00	1,00	6.44	0.011	1,00	0.019
	s(psc)	1,00	1,00	1.39	0.238	1,00	0.346
	s(osc)	1,00	1,00	1.41	0.237	1,00	0.346
	ti(WordFrequency,Time)	17.56	17.97	123.07	<0.001	<0.001	<0.001
	ti(NumberOfLetters,Time)	17.34	17.95	91.72	<0.001	<0.001	<0.001
	ti(psc,Time)	14.12	15.33	16.11	<0.001	<0.001	<0.001
	ti(osc,Time)	14.93	15.74	33.47	<0.001	<0.001	<0.001
	ti(psc,osc)	1.03	1.03	0.7	0.406	1,00	0.533
	ti(psc,osc,Time)	30.78	33.88	19.52	<0.001	<0.001	<0.001
	s(WORD)	678.04	683,00	138.63	<0.001	<0.001	<0.001
FC6	s(WordFrequency)	1,00	1,00	0.62	0.433	1,00	0.56
	s(Time)	9,00	9,00	16640.15	<0.001	<0.001	<0.001
	s(NumberOfLetters)	1,00	1,00	4.01	0.045	1,00	0.074
	s(psc)	1,00	1,00	0.81	0.366	1,00	0.492
	s(osc)	1,00	1,00	0.33	0.566	1,00	0.697
	ti(WordFrequency,Time)	17.39	17.95	86.61	<0.001	<0.001	<0.001

	ti(NumberOfLetters,Time)	17.39	17.96	90.35	<0.001	<0.001	<0.001
	ti(psc,Time)	14.68	15.62	20.66	<0.001	<0.001	<0.001
	ti(osc,Time)	14.2	15.32	27.48	<0.001	<0.001	<0.001
	ti(psc,osc)	1.02	1.02	0.19	0.67	1,00	0.791
	ti(psc,osc,Time)	31.7	34.46	18.45	<0.001	<0.001	<0.001
	s(WORD)	678.45	683,00	151.08	<0.001	<0.001	<0.001
P4	s(WordFrequency)	1,00	1,00	0.59	0.442	1,00	0.57
	s(Time)	9,00	9,00	5531.4	<0.001	<0.001	<0.001
	s(NumberOfLetters)	1.02	1.02	4.08	0.045	1,00	0.074
	s(psc)	1,00	1,00	0.11	0.74	1,00	0.836
	s(osc)	1,00	1,00	0.04	0.848	1,00	0.905
	ti(WordFrequency,Time)	17.02	17.88	59.75	<0.001	<0.001	<0.001
	ti(NumberOfLetters,Time)	17.39	17.95	141.08	<0.001	<0.001	<0.001
	ti(psc,Time)	15.42	15.93	22.25	<0.001	<0.001	<0.001
	ti(osc,Time)	15.46	15.93	32.07	<0.001	<0.001	<0.001
	ti(psc,osc)	1.03	1.03	0.31	0.582	1,00	0.709
	ti(psc,osc,Time)	27.8	31.57	13.71	<0.001	<0.001	<0.001

	s(WORD)	677.71	683,00	129.69	<0.001	<0.001	<0.001
CP2	s(WordFrequency)	1,00	1,00	0.28	0.597	1,00	0.724
	s(Time)	9,00	9,00	15677.93	<0.001	<0.001	<0.001
	s(NumberOfLetters)	1.02	1.02	3.94	0.048	1,00	0.078
	s(psc)	1,00	1,00	0.13	0.72	1,00	0.828
	s(osc)	1,00	1,00	0.07	0.796	1,00	0.866
	ti(WordFrequency,Time)	16.69	17.78	43.55	<0.001	<0.001	<0.001
	ti(NumberOfLetters,Time)	17.37	17.95	140.56	<0.001	<0.001	<0.001
	ti(psc,Time)	15.22	15.88	20.38	<0.001	<0.001	<0.001
	ti(osc,Time)	15.2	15.85	34.89	<0.001	<0.001	<0.001
	ti(psc,osc)	1.02	1.02	0.03	0.872	1,00	0.925
	ti(psc,osc,Time)	31.1	34.37	10.56	<0.001	<0.001	<0.001
	s(WORD)	678.5	683,00	153.08	<0.001	<0.001	<0.001
F8	s(WordFrequency)	1,00	1,00	0.98	0.321	1,00	0.435
	s(Time)	9,00	9,00	16874.28	<0.001	<0.001	<0.001
	s(NumberOfLetters)	1,00	1,00	2.95	0.086	1,00	0.135
	s(psc)	1,00	1,00	0.31	0.579	1,00	0.709

	s(osc)	1,00	1,00	1.58	0.208	1,00	0.306
	ti(WordFrequency,Time)	17.42	17.96	107.59	<0.001	<0.001	<0.001
	ti(NumberOfLetters,Time)	17.34	17.95	60.92	<0.001	<0.001	<0.001
	ti(psc,Time)	14.76	15.67	21.54	<0.001	<0.001	<0.001
	ti(osc,Time)	14,00	15.29	23.47	<0.001	<0.001	<0.001
	ti(psc,osc)	1.06	1.06	0,00	0.963	1,00	0.983
	ti(psc,osc,Time)	31.47	34.36	17.56	<0.001	<0.001	<0.001
	s(WORD)	678.55	683,00	155.71	<0.001	<0.001	<0.001
F7	s(WordFrequency)	1.02	1.02	0.08	0.766	1,00	0.854
	s(Time)	9,00	9,00	22258.8	<0.001	<0.001	<0.001
	s(NumberOfLetters)	1,00	1,00	3.77	0.052	1,00	0.084
	s(psc)	1,00	1,00	0.11	0.745	1,00	0.839
	s(osc)	1,00	1,00	0.15	0.695	1,00	0.812
	ti(WordFrequency,Time)	17.42	17.95	113.98	<0.001	<0.001	<0.001
	ti(NumberOfLetters,Time)	17.57	17.98	83.7	<0.001	<0.001	<0.001
	ti(psc,Time)	14.62	15.61	19.33	<0.001	<0.001	<0.001
	ti(osc,Time)	14.93	15.77	35.51	<0.001	<0.001	<0.001

	ti(psc,osc)	1.1	1.1	0.97	0.37	1,00	0.493
	ti(psc,osc,Time)	33.11	35.4	28.18	<0.001	<0.001	<0.001
	s(WORD)	678.47	683,00	154.5	<0.001	<0.001	<0.001
Oz	s(WordFrequency)	1.01	1.01	0.02	0.885	1,00	0.931
	s(Time)	8.99	9,00	4708.52	<0.001	<0.001	<0.001
	s(NumberOfLetters)	1.01	1.01	6.83	0.009	1,00	0.015
	s(psc)	1,00	1,00	1.33	0.249	1,00	0.358
	s(osc)	1,00	1,00	0.19	0.663	1,00	0.784
	ti(WordFrequency,Time)	17.25	17.93	80.61	<0.001	<0.001	<0.001
	ti(NumberOfLetters,Time)	16.65	17.79	93.73	<0.001	<0.001	<0.001
	ti(psc,Time)	15.09	15.82	24.71	<0.001	<0.001	<0.001
	ti(osc,Time)	14.93	15.74	22.36	<0.001	<0.001	<0.001
	ti(psc,osc)	1.04	1.04	0.82	0.368	1,00	0.492
	ti(psc,osc,Time)	28.53	32.49	10.11	<0.001	<0.001	<0.001
	s(WORD)	675.53	683,00	91.56	<0.001	<0.001	<0.001
FC5	s(WordFrequency)	1.01	1.01	1.74	0.185	1,00	0.277
	s(Time)	9,00	9,00	20878.18	<0.001	<0.001	<0.001

	s(NumberOfLetters)	1,00	1,00	2.31	0.128	1,00	0.195
	s(psc)	1,00	1,00	0.53	0.464	1,00	0.594
	s(osc)	1,00	1,00	0,00	0.996	1,00	0.996
	ti(WordFrequency,Time)	17.53	17.97	84.24	<0.001	<0.001	<0.001
	ti(NumberOfLetters,Time)	17.42	17.96	89.79	<0.001	<0.001	<0.001
	ti(psc,Time)	14.49	15.56	14.19	<0.001	<0.001	<0.001
	ti(osc,Time)	14.66	15.62	37.99	<0.001	<0.001	<0.001
	ti(psc,osc)	1.03	1.03	0.43	0.526	1,00	0.661
	ti(psc,osc,Time)	32.17	34.96	21.89	<0.001	<0.001	<0.001
	s(WORD)	678.29	683,00	146.02	<0.001	<0.001	<0.001
Fpz	s(WordFrequency)	1,00	1,00	3.18	0.074	1,00	0.118
	s(Time)	9,00	9,00	18818.55	<0.001	<0.001	<0.001
	s(NumberOfLetters)	1,00	1,00	5.41	0.02	1,00	0.033
	s(psc)	1,00	1,00	0.44	0.509	1,00	0.642
	s(osc)	1,00	1,00	3.17	0.076	1,00	0.119
	ti(WordFrequency,Time)	17.57	17.98	158.09	<0.001	<0.001	<0.001
	ti(NumberOfLetters,Time)	17.28	17.94	71.12	<0.001	<0.001	<0.001

	ti(psc,Time)	14.43	15.61	14.66	<0.001	<0.001	<0.001
	ti(osc,Time)	15.11	15.8	47.01	<0.001	<0.001	<0.001
	ti(psc,osc)	1.39	1.39	0.21	0.782	1,00	0.859
	ti(psc,osc,Time)	32,00	34.65	28.84	<0.001	<0.001	<0.001
	s(WORD)	678.82	683,00	178.99	<0.001	<0.001	<0.001
O1	s(WordFrequency)	1,00	1,00	0,00	0.993	1,00	0.996
	s(Time)	9,00	9,00	11284.97	<0.001	<0.001	<0.001
	s(NumberOfLetters)	1.03	1.03	0.83	0.381	1,00	0.505
	s(psc)	1,00	1,00	0.04	0.838	1,00	0.9
	s(osc)	1,00	1,00	0.02	0.874	1,00	0.925
	ti(WordFrequency,Time)	17.47	17.96	132.06	<0.001	<0.001	<0.001
	ti(NumberOfLetters,Time)	17.28	17.94	58.92	<0.001	<0.001	<0.001
	ti(psc,Time)	15.31	15.9	18.45	<0.001	<0.001	<0.001
	ti(osc,Time)	14.99	15.76	23.25	<0.001	<0.001	<0.001
	ti(psc,osc)	1.02	1.02	0,00	0.957	1,00	0.979
	ti(psc,osc,Time)	31.59	34.75	15.17	<0.001	<0.001	<0.001
	s(WORD)	675.96	683,00	97.12	<0.001	<0.001	<0.001

Table A3. To justify the use of a three-way interaction between time, OSC and PSC in the models used in the analysis (full models), we ran a simplified version of such models (simple models), not including the interaction of interest. We then compared both models' fit to the data by means of the Akaike Information Criterion (AIC). For all electrodes, the AIC of the full model was smaller than the AIC of the simple one, meaning that the inclusion of the three-way interaction improved the fit. The table shows results of the model comparison. The first column reports the electrode names. The second and third column represent the AIC for the models used in the analysis (full model) and the simplified one (simple model). The fourth column represent the difference between the AIC of the full and the simple models.

Electrode	**AIC full model**	**AIC simple model**	**AIC difference**
T8	538939.49	539511.95	-572.46
CP1	610297.19	610523.73	-226.54
CP6	573308.27	573740.02	-431.75
C3	583443.92	583834.2	-390.28
Cz	650101.06	650503.55	-402.49
C4	618256.41	618589.2	-332.79
P7	439040.03	439525.34	-485.31
Pz	641645.63	642037.37	-391.74

P3	573014.05	573353.37	-339.32
T7	484714.19	485812.92	-1098.73
Fp2	644742.73	646036.95	-1294.22
FC2	637309.73	637846.86	-537.13
Fp1	623615.11	624553.24	-938.13
CP5	502205.68	502518.25	-312.57
O2	659015.17	659445.64	-430.47
P8	552052.39	552784.02	-731.63
F3	613711.95	614460.46	-748.51
Fz	641073.56	641665.94	-592.38
FC1	632134.11	632637.78	-503.67
F4	618864.86	619500.6	-635.74
FC6	576844.1	577448.22	-604.12
P4	616210.26	616624.49	-414.23
CP2	627412.84	627748.01	-335.17
F8	568629.99	569199.96	-569.97
F7	529915.78	530885.83	-970.05

Oz	608948.16	609251.25	-303.09
FC5	551247.69	551985.12	-737.43
Fpz	622312.89	623284.35	-971.46
O1	625214.61	625717.43	-502.82

Figure A1. ERP grandaverage of Kilo-word stimuli. The figure shows the grandaverage including all words of Kilo-word database used for the analysis. Only words in which both OSC and PSC were equal to 1 were excluded from the initial set.

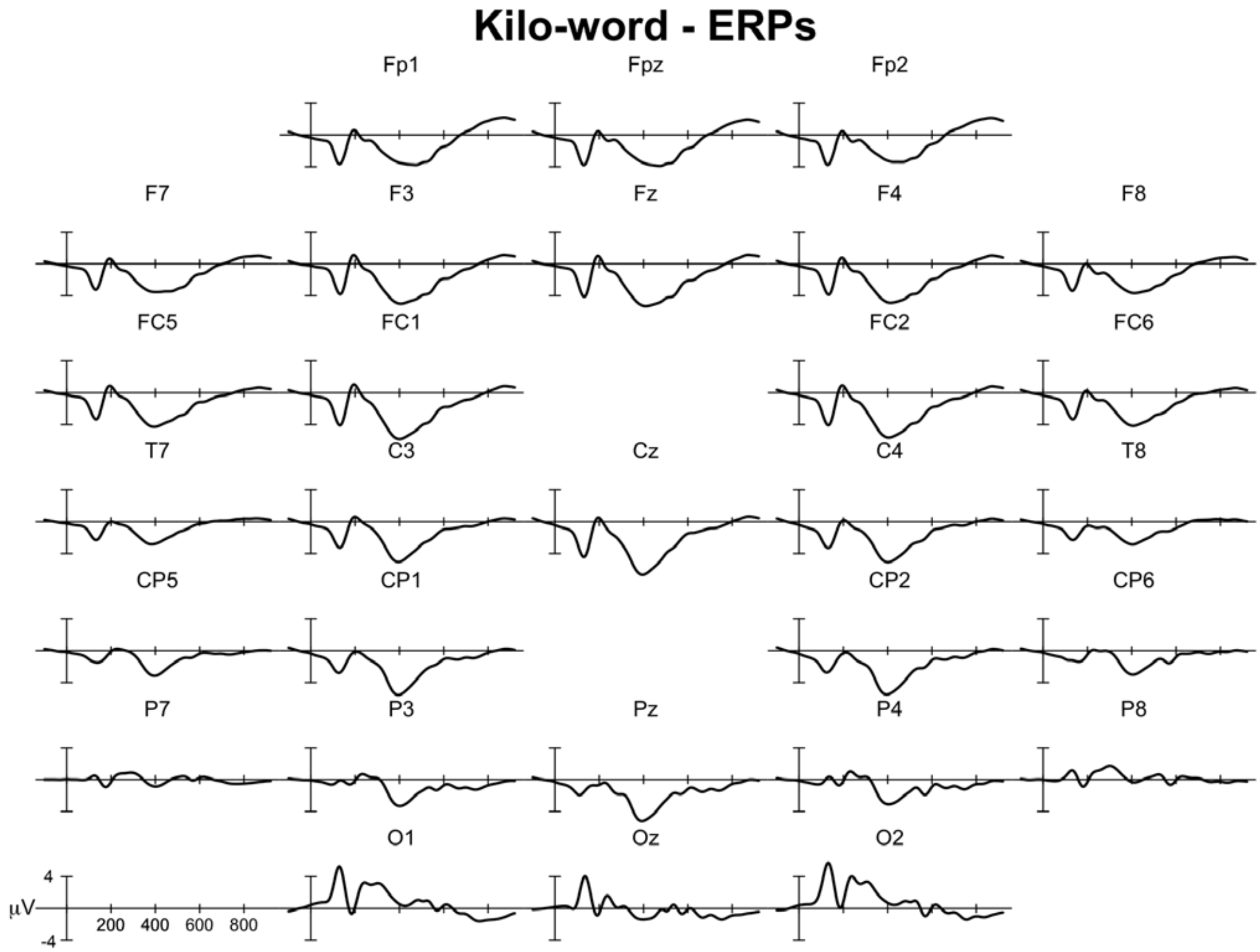


Figure A2. Scalp plot of partial effects of the interaction between Length and Time. The contour map for each electrode is a topographic representation of the partial effect of Time (in the x-axis) and Length (in the y-axis). Amplitude is codified as color using a jet palette: color towards red indicates positive values, while color towards blue indicates negative values, while colors toward green indicate in-between values. Topographic maps have been masked so that only effect estimates whose confidence interval at 95% did not include zero were included. Electrodes are reported in a regular grid array that approximate their position on the scalp.

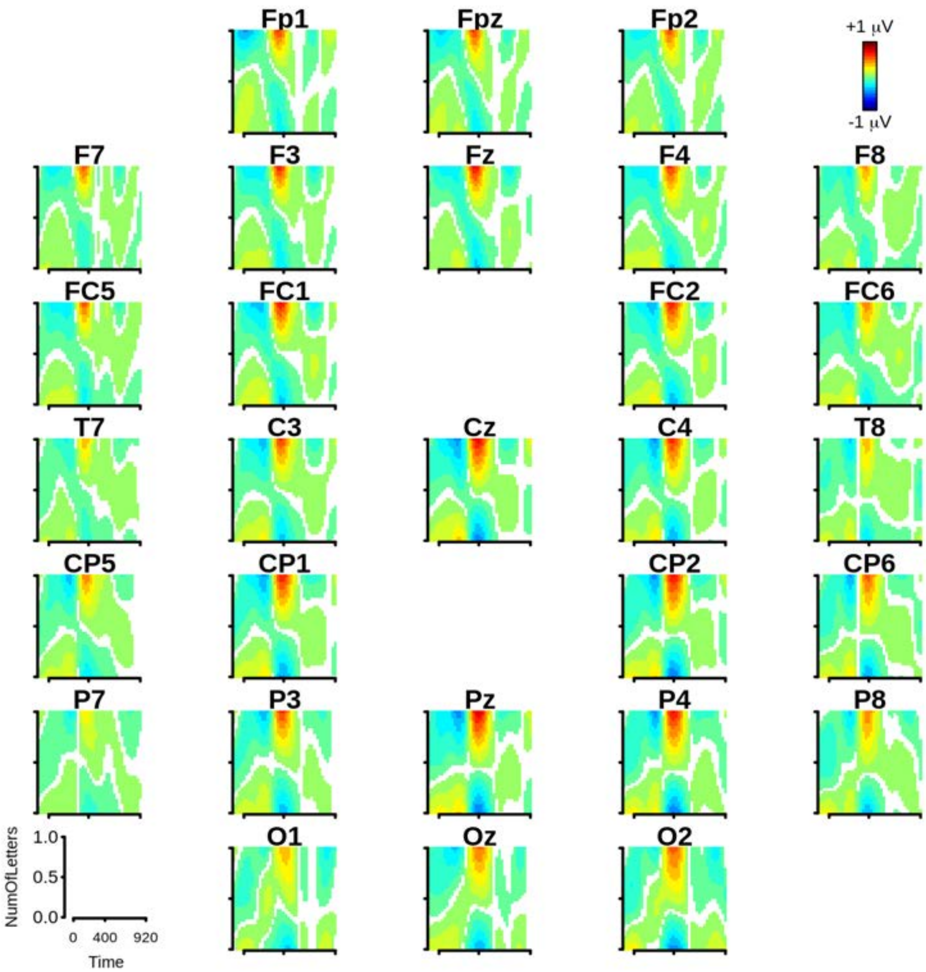
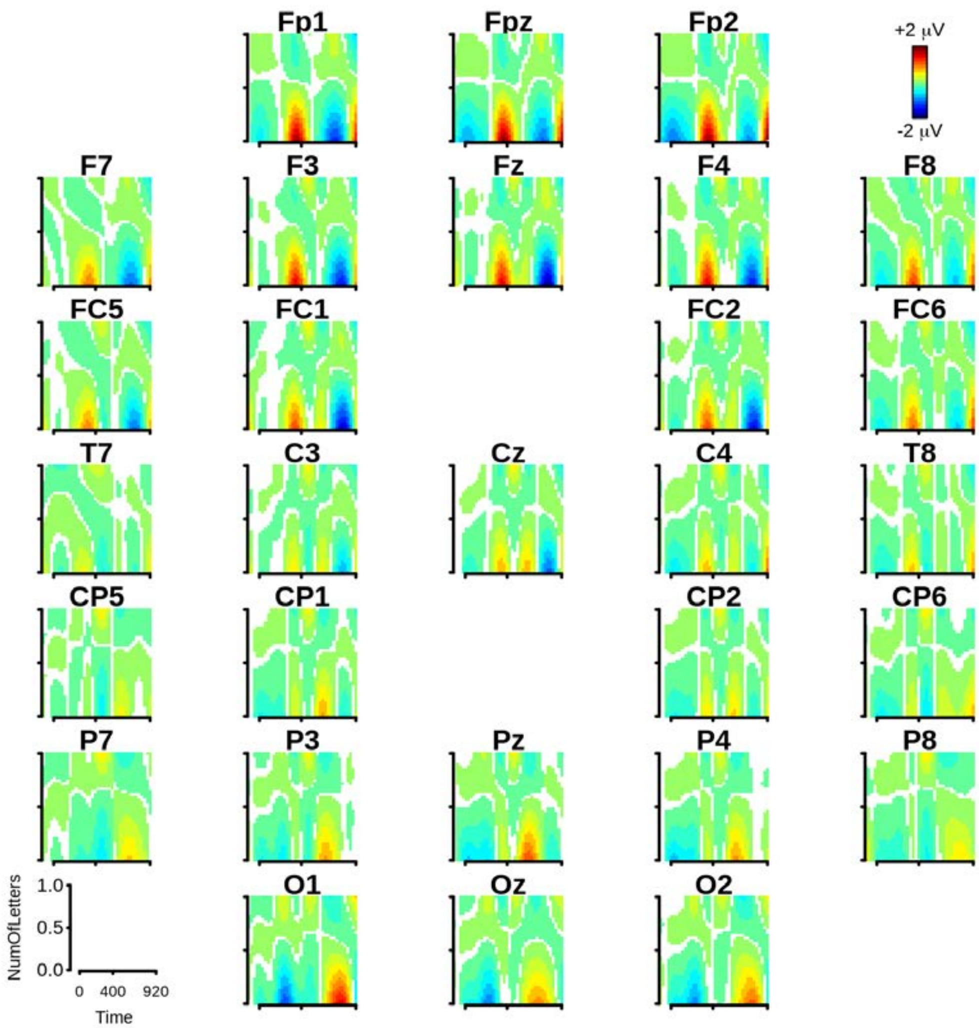
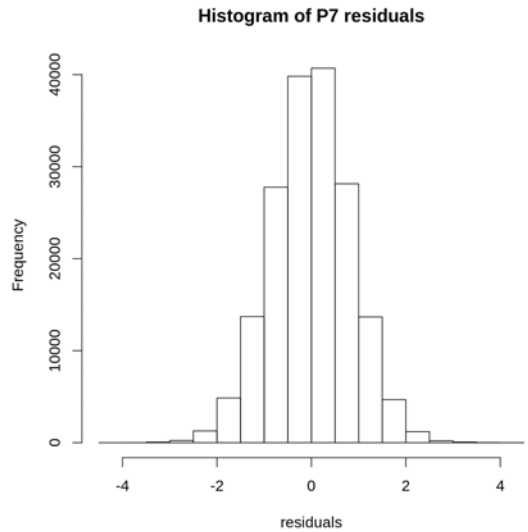
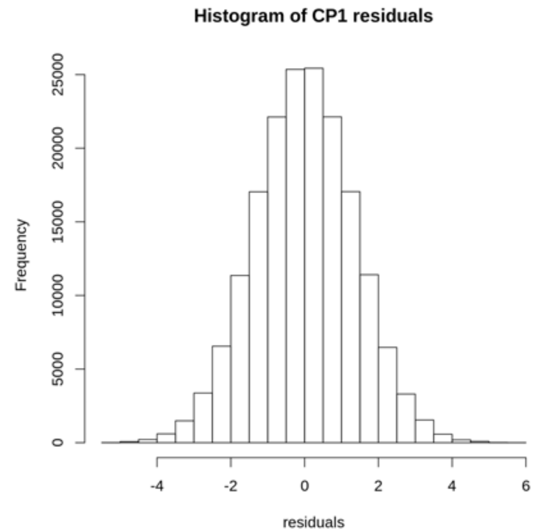


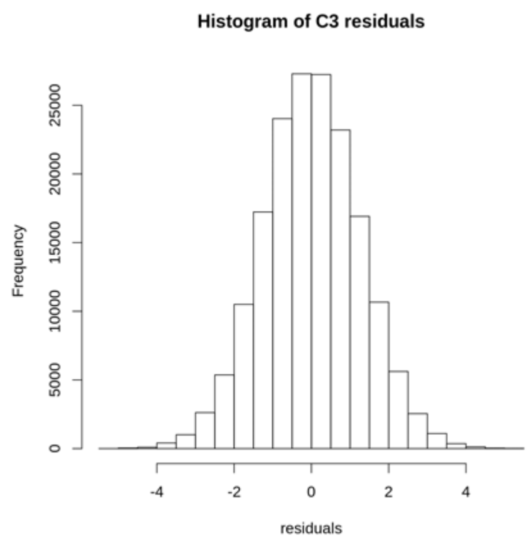
Figure A3. Scalp plot of partial effects of the interaction between Frequency and Time. The contour map for each electrode is a topographic representation of the partial effect of Time (in the x-axis) and Frequency (in the y-axis). Amplitude is codified as color using a jet palette: color towards red indicates positive values, while color towards blue indicates negative values, while colors toward green indicate in-between values. Topographic maps have been masked so that only effect estimates whose confidence interval at 95% did not include zero were included. Electrodes are reported in a regular grid array that approximate their position on the scalp.

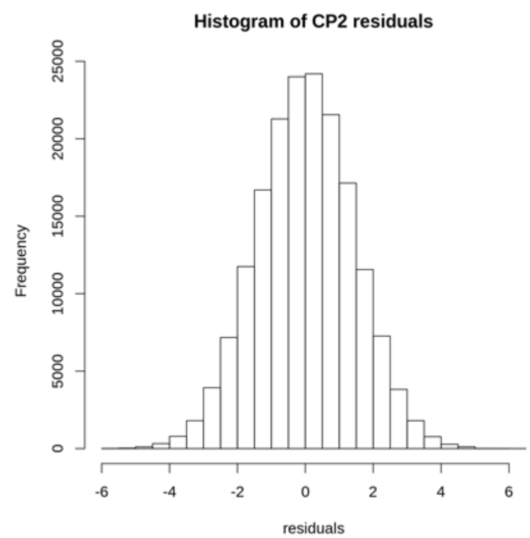


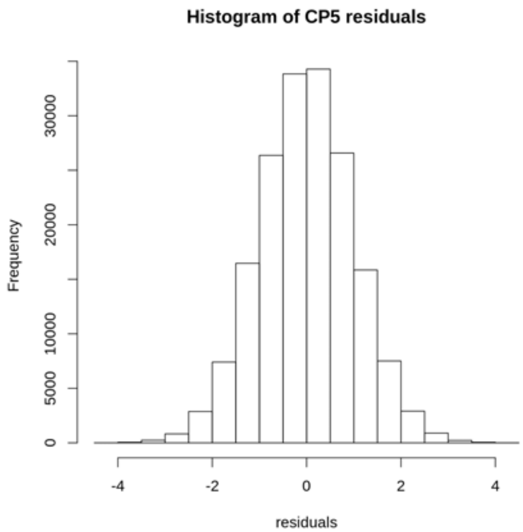
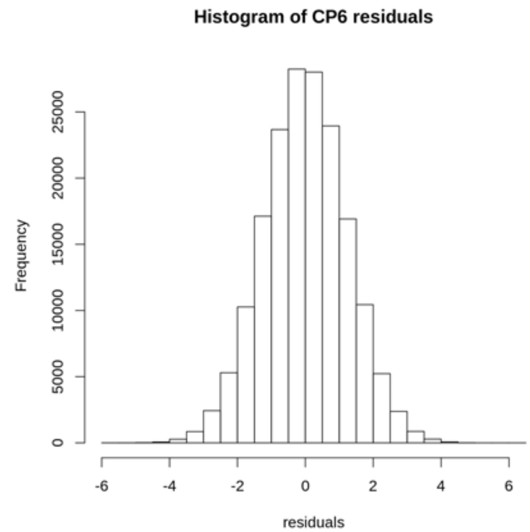
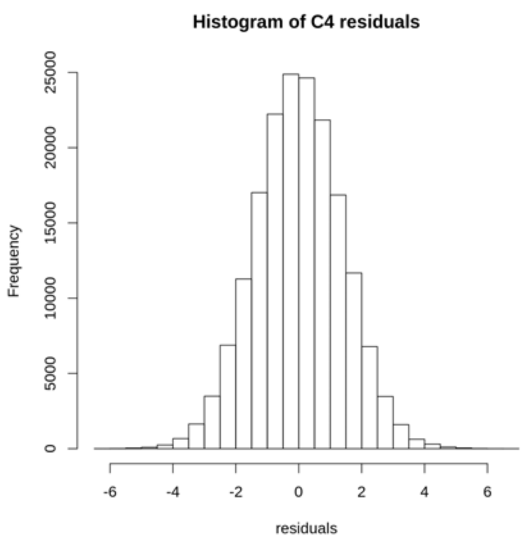
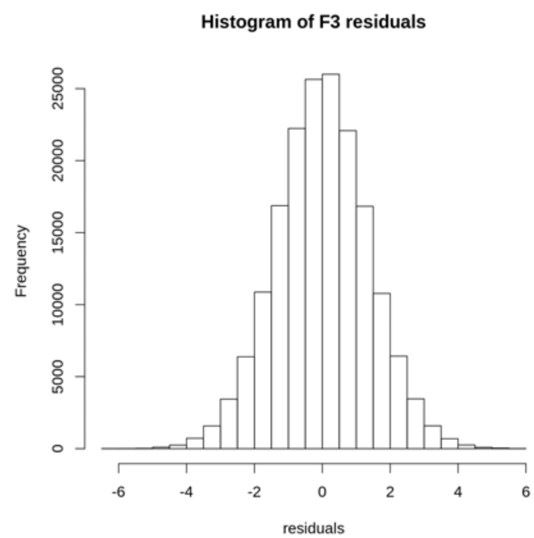
Figures A4 - A33. Histograms of residuals for each electrode's model. Residuals of all the models were normally distributed.

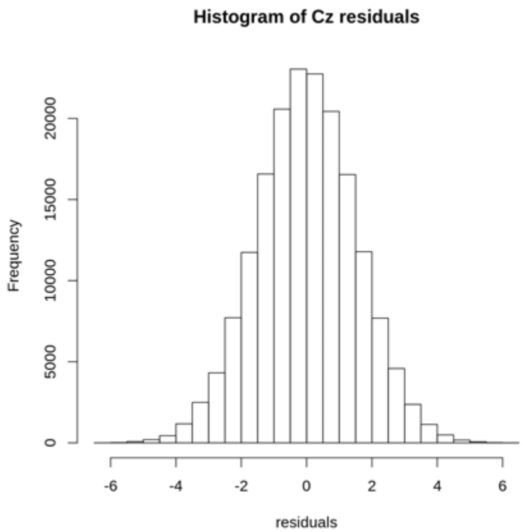


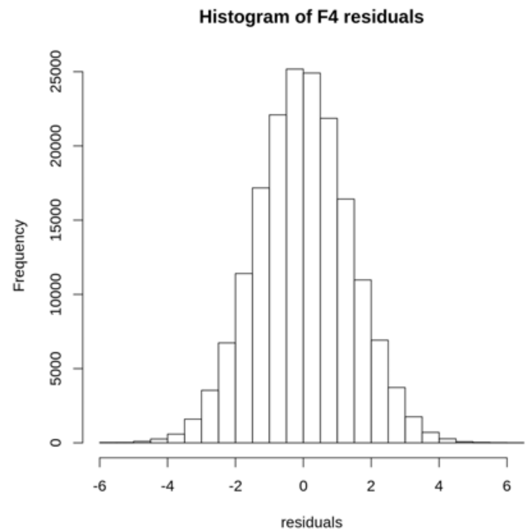


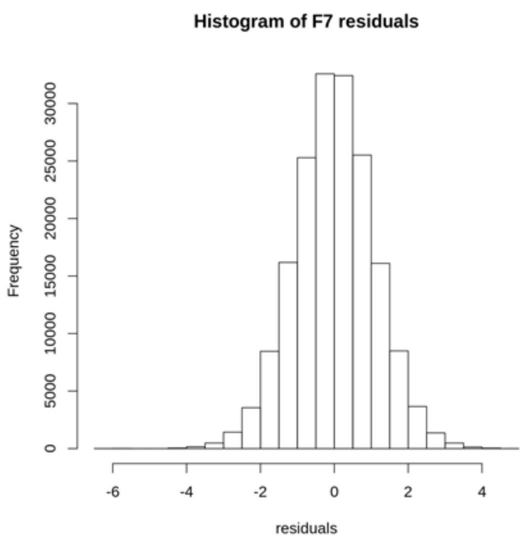


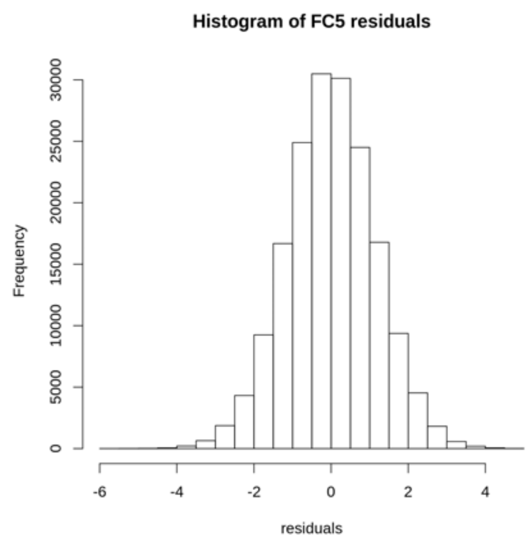


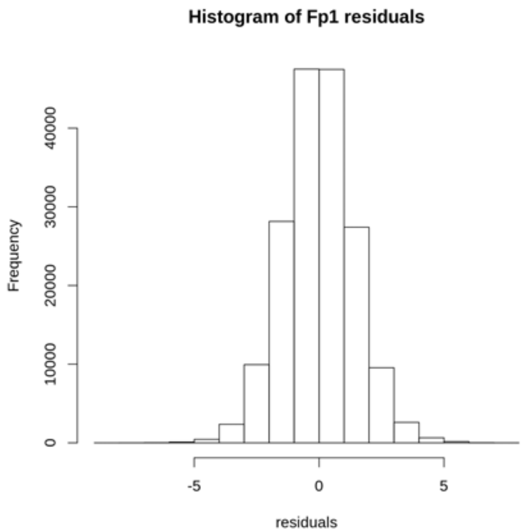

Histogram of CP5 residuals
Frequency
0
5000
10000
20000
30000
-4
-2
0
2
4
residuals

Histogram of CP6 residuals
Frequency
0
5000
10000
15000
20000
25000
-6
-4
-2
0
2
4
6
residuals

Histogram of C4 residuals
Frequency
0
5000
10000
15000
20000
25000
-6
-4
-2
0
2
4
6
residuals

Histogram of F3 residuals
Frequency
0
5000
10000
15000
20000
25000
-6
-4
-2
0
2
4
6
residuals

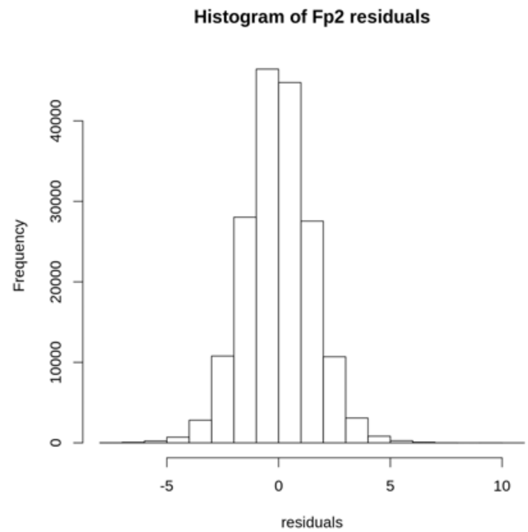

Histogram of Cz residuals
Frequency
0
5000
10000
15000
20000
-6
-4
-2
0
2
4
6
residuals

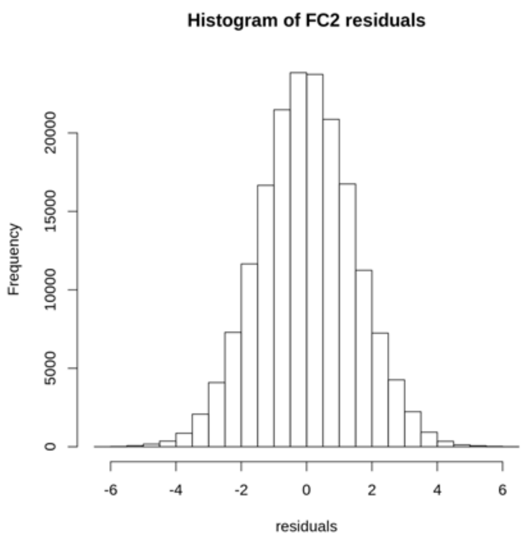

Histogram of F4 residuals
Frequency
0
5000
10000
15000
20000
25000
-6
-4
-2
0
2
4
6
residuals

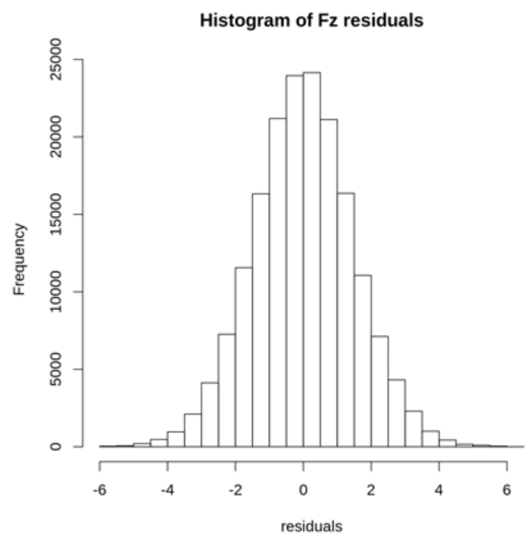

Histogram of F7 residuals
Frequency
0
5000
10000
15000
20000
25000
30000
-6
-4
-2
0
2
4
residuals

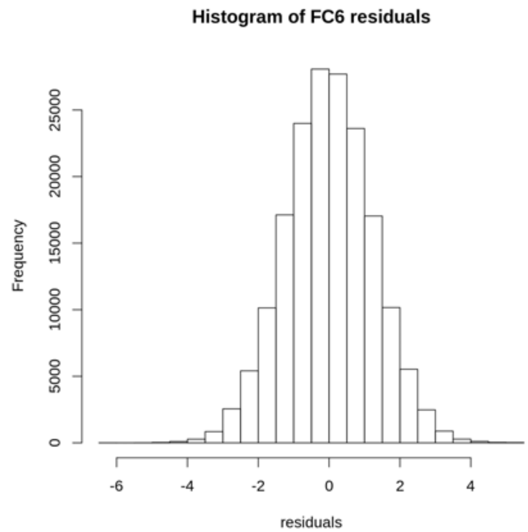

Histogram of FC5 residuals
Frequency
0
5000
10000
15000
20000
25000
30000
-6
-4
-2
0
2
4
residuals

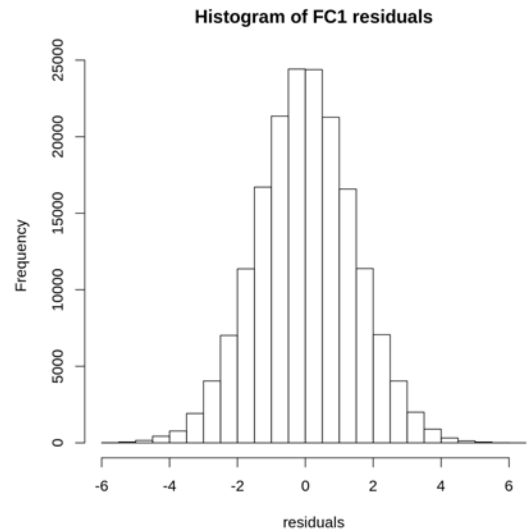

Histogram of Fp1 residuals
Frequency
0
10000
20000
30000
40000
-5
0
5
residuals

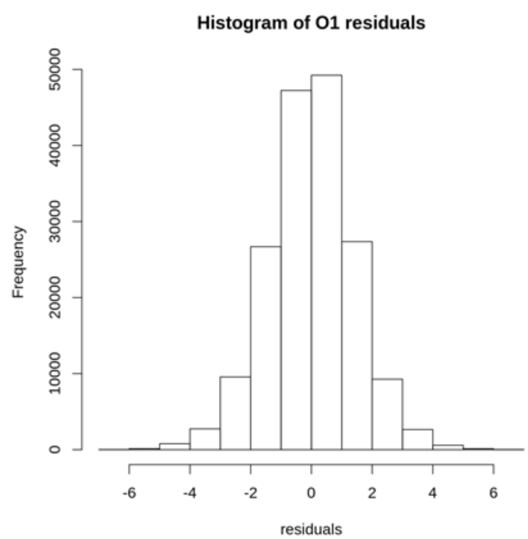

Histogram of Fp2 residuals
Frequency
0
10000
20000
30000
40000
-5
0
5
10
residuals

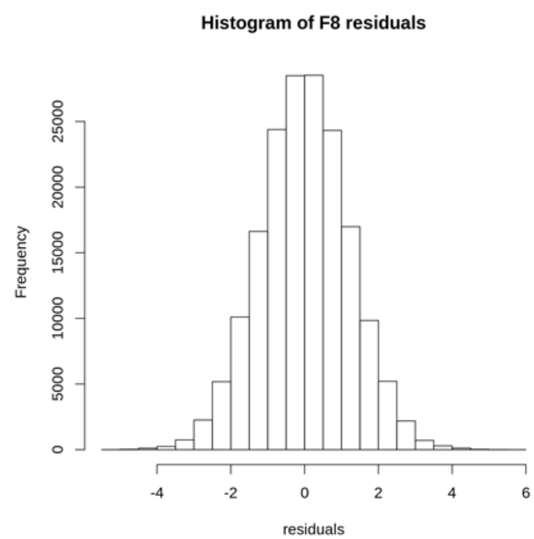

Histogram of FC2 residuals
Frequency
0
5000
10000
15000
20000
-6
-4
-2
0
2
4
6
residuals

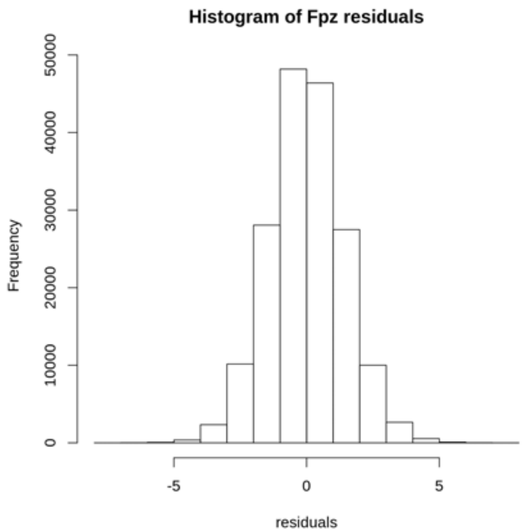

Histogram of Fz residuals
Frequency
0
5000
10000
15000
20000
25000
-6
-4
-2
0
2
4
6
residuals

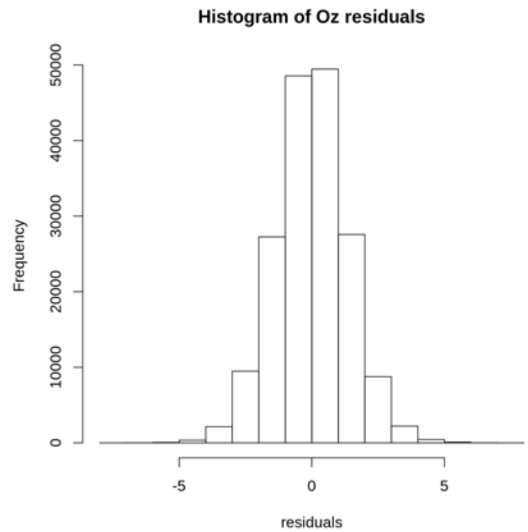

Histogram of FC6 residuals
Frequency
0
5000
10000
15000
20000
25000
-6
-4
-2
0
2
4
residuals

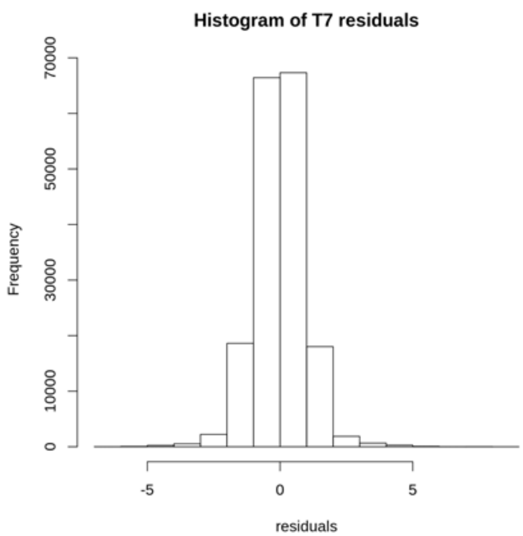

Histogram of FC1 residuals
Frequency
0
5000
10000
15000
20000
25000
-6
-4
-2
0
2
4
6
residuals

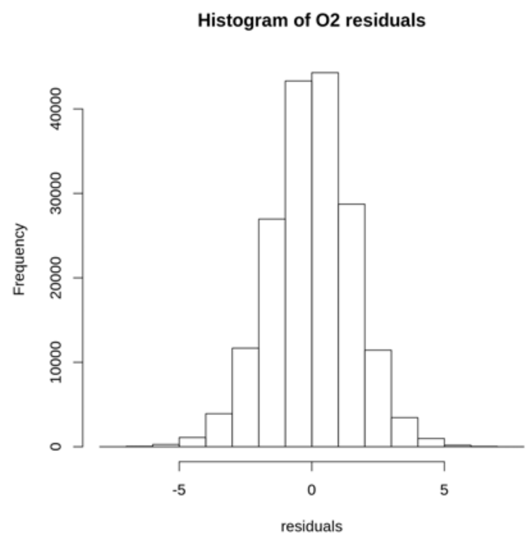

Histogram of O1 residuals
Frequency
0
10000
20000
30000
40000
50000
-6
-4
-2
0
2
4
6
residuals

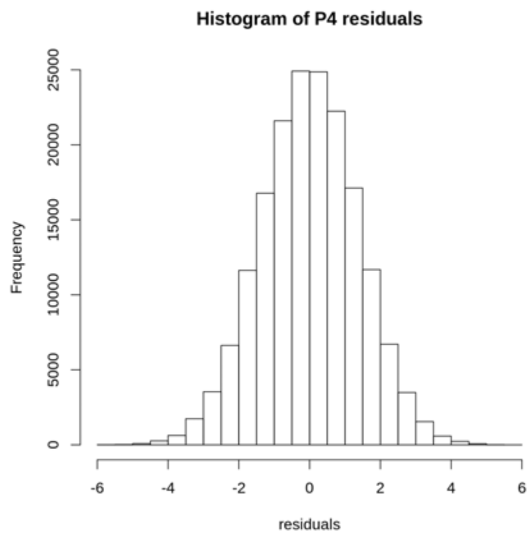

Histogram of F8 residuals
Frequency
0
5000
10000
15000
20000
25000
-4
-2
0
2
4
6
residuals

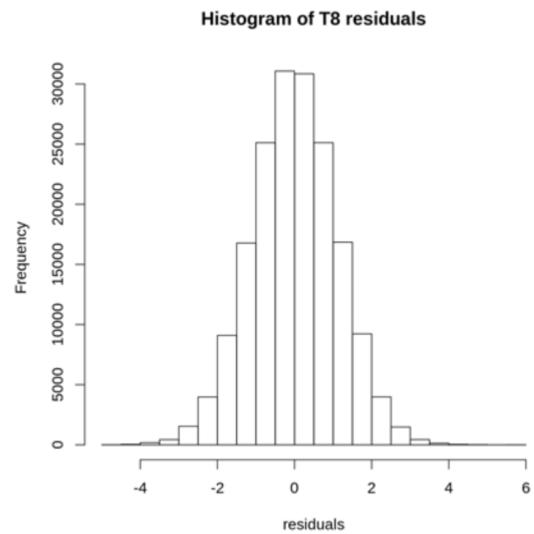

Histogram of Fpz residuals
Frequency
0
10000
20000
30000
40000
50000
-5
0
5
residuals

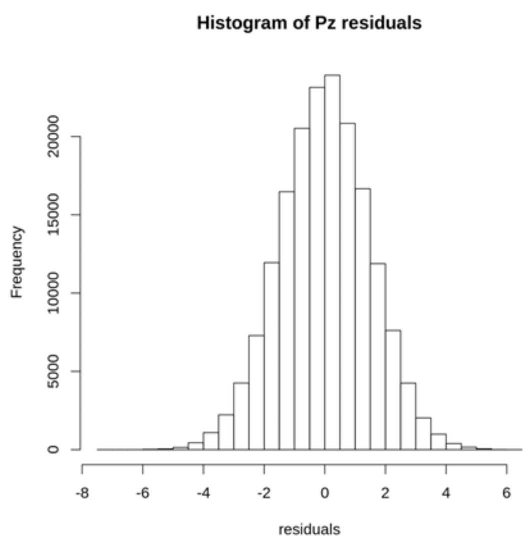

Histogram of Oz residuals
Frequency
0
10000
20000
30000
40000
50000
-5
0
5
residuals

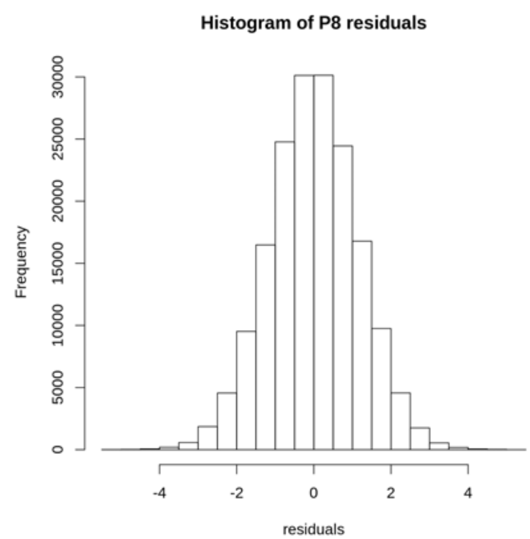

Histogram of T7 residuals
Frequency
0
10000
30000
50000
70000
-5
0
5
residuals

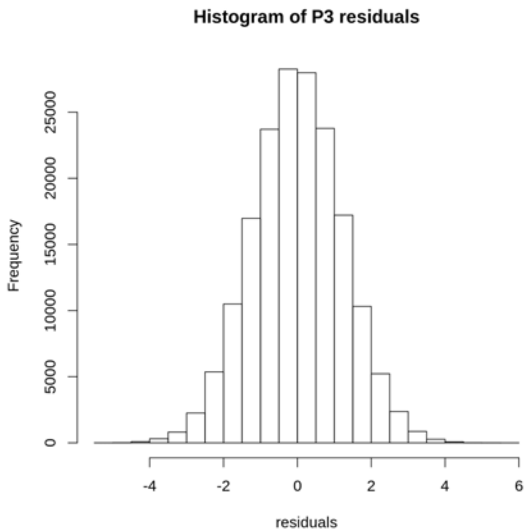

Histogram of O2 residuals
Frequency
0
10000
20000
30000
40000
-5
0
5
residuals


Histogram of P4 residuals
Frequency
0
5000
10000
15000
20000
25000
-6
-4
-2
0
2
4
6
residuals


Histogram of T8 residuals
Frequency
0
5000
10000
15000
20000
25000
30000
-4
-2
0
2
4
6
residuals


Histogram of Pz residuals
Frequency
0
5000
10000
15000
20000
-8
-6
-4
-2
0
2
4
6
residuals


Histogram of P8 residuals
Frequency
0
5000
10000
15000
20000
25000
30000
-4
-2
0
2
4
residuals


Histogram of P3 residuals
Frequency
0
5000
10000
15000
20000
25000
-4
-2
0
2
4
6
residuals