Using Deconvolution to Understand Cognitive Processes in Reading and Language Comprehension

Jeremiah Adeyemi

Supervised by: Dr. Sara Milligan and Dr. Elizabeth R. Schotter, Eye Movements and Cognition Lab, University of South Florida

Final Report for CAS Undergraduate Research Fellowship in Visualization, Virtualization, and Simulation (VVS)

4th March, 2025

Project Overview

Fewer recent developments in the cognitive science of reading exhibit greater potential than co-registration, an approach that simultaneously measures eye movement behaviors via eye tracking and neural responses via the electroencephalogram (EEG) to characterize a person's natural reading behavior. Up to the last decade, these two fields have yielded separate insights into the nature of comprehension during reading under experimentally controlled conditions. For instance, EEG researchers who study language processing have traditionally employed Rapid Serial Visual Presentation (RSVP) - a procedure in which participants are shown words one at a time, generally presented in the middle of a computer screen (Degno et. al, 2019). While this has its benefits in helping us study isolated neural responses, the field recognizes that our insights from such methods may not be confidently generalized as natural reading shows more complex dynamics. RSVP is a relatively unnatural reading experience and also does not allow the reader to control the sequence of which words they look at or how long they decide to look at them.

An alternative to the RSVP paradigm is to present entire sentences on the screen and allow readers to freely move their eyes while recording the EEG signal and eye movements, using an eye tracking camera. This allows for a more ecologically valid reading experience and testing how readers process language in a more natural scenario. Data from the eye movements and neural recording can then be synchronized (i.e., co-registration of EEG and eye tracking data) to extract neural responses to words and linguistic manipulations of interest. That being said, this new method comes with its challenges, as we have no control over our participants' reading behaviors. The eye movements people make during reading are quick - around 200 ms (Rayner, 2009), and this presents us with unique challenges. First, the EEG responses have to be accurately synchronized to eye movement data. Second, neural overlaps often occur from subsequent fixations (a participant's gaze on a word). Third, it is important to control for covariates and confounds that can impact our results, such as characteristics of the words influencing the timing between fixations of interest.

One of the challenges in analyzing effects related to language processing during reading in a natural reading co-registration paradigm (i.e., simultaneous recording of EEG and eye tracking) is that one of the most extensively studied EEG responses, the N400 ERP component tends to peak in amplitude around 400 milliseconds after an eye fixation begins (see Kutas & Federmeier, 2014). Therefore, this response to a particular word tends to occur after the reader has already moved their eyes to the next word, and therefore neural activity in response to the subsequently fixated word may contaminate the neural responses measured on the word of interest. Therefore, it is important to evaluate the validity of the responses we are measuring to be confident that this temporal overlap is not producing spurious effects in the data.

While methods have been developed to solve these issues using advanced tools and computation, the second one possesses the most impact. In particular, the fixation-related responses can not be compared directly with each other, as their baselines vary. Moreover, neural responses from different fixations can also overlap with responses from the initial stimuli presented (initial word view), which undulates longer over time (Dimigen & Ehinger, 2021).

These neural overlaps result from temporal variations in fixating over time, which, till this research, our lab has ascribed to the natural variance that arises in our co-registration studies. Our question: to what degree can we reduce this variance, and how significantly do our results change from doing so? To answer this, we turn to a tool that has been developed to advance the intersection of mathematics and psychology: deconvolution modeling.

Observed EEG signals are a linear sum of brain activity in the regions studied, represented by voltage per unit time. For neural responses that happen in quick succession, this sum happens to be a **convolution** (see Figure 1 for example diagram) - each neural event extends over time, and what we see at any moment is the brain's true activity summed with the lingering effect caused by the prior event. In the diagram below, stimulus responses are more long-lasting, overlapping with more occurrent and short-lived fixation responses. Previous experimental approaches would manually add time between these responses by presenting words serially, but this overlap becomes challenging in natural reading scenarios.

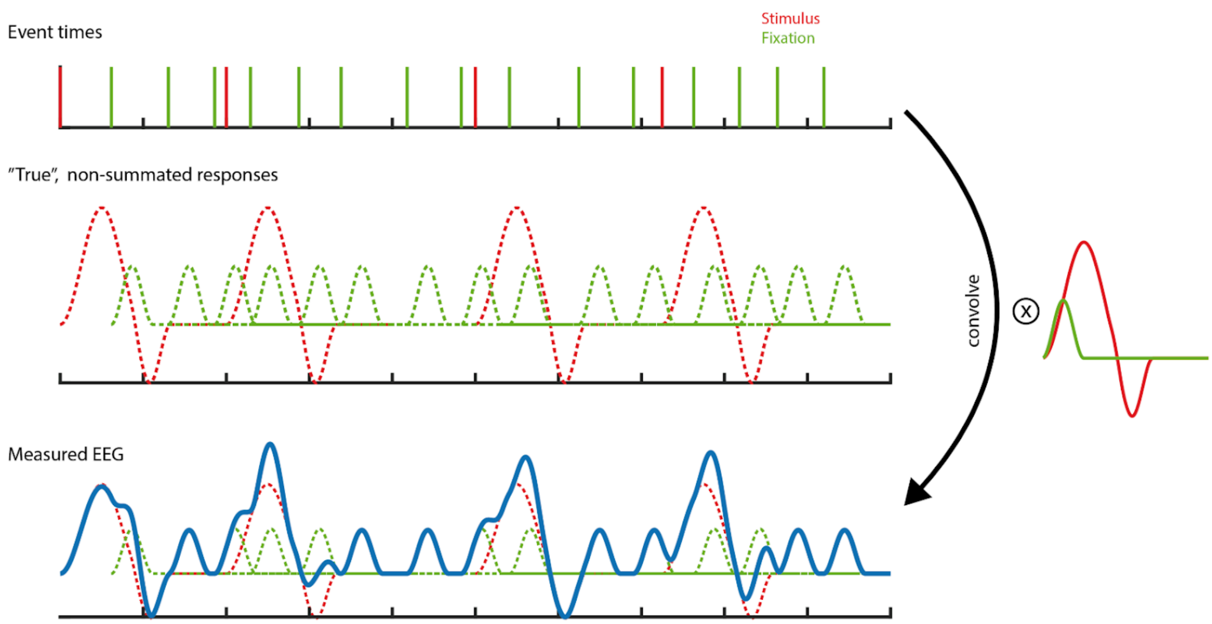

Figure 1. Diagram of convolution (taken from unfoldtoolbox.org, Ehinger & Dimigen 2018)

Because this overlap is a mathematical function, it follows that to isolate the neural effects we are interested in, we apply an inverse function to these sums, also known as **deconvolution**. Many approaches to this method exist, each with varying accuracies. Our method of choice is *time-expanded linear deconvolution* – a linear regression process that models the course of neural signals over time, based on a set of predictor variables.

With this method, we can define the events of interest, express the EEG signal as a summation of likely effects and interactions *(predictors earlier mentioned)* within some time interval, and solve for the coefficients of those predictors, thereby quantifying their impact on the signal. In practice, these regression coefficients (betas) are solved by multiplication with a design matrix; its rows correspond to time points of interest in the EEG, and its columns follow predictors at those times and the times surrounding them (described as local time - sampled time points after event onset). These betas are non-overlapping and can then be plotted in time, allowing us to see overall neural responses and the specific effects our predictors have when dissociated from one another (Dimigen & Ehinger, 2021). This method relies on the fact that the event overlaps in a signal are temporally unique, allowing us to expect unique estimates as coefficients. We can further bolster this by the variation in our experiments.

To apply this, we use the *unfold* toolbox (Dimigen & Ehinger, 2021), an open-source MATLAB repository that works as described above. It also allows us to factor in non-linear predictors with spline regression (linear estimates for non-linear effects), use a mass-univariate linear modeling method (helping us see predictor effects without deconvolution), apply baseline correction to signals, and perform data visualizations. We use this method as part of a larger study on refixations: quick eye movements made towards a previously viewed position in a text.

Current study rationale and data overview

The current study was designed to assess the relative neural responses to separate types of fixations on words. Occasionally readers will look at a word multiple times before moving on (i.e., refixate; McConkie et al., 1989), and we aimed to use co-registration to better understand if these refixations are essentially random or if they serve some purpose in achieving language comprehension. People tend to refixate words more often when they are longer words, so in our design, we presented sentences with particular *target words* of interest that were between 7 - 13 characters long to increase the chances that people would refixate. We also included a manipulation of that word's expectancy in the sentence, which has been shown to influence the N400 response (i.e., a more negative amplitude of the electrical brain signal when words are unexpected compared to expected; see Figure 2).

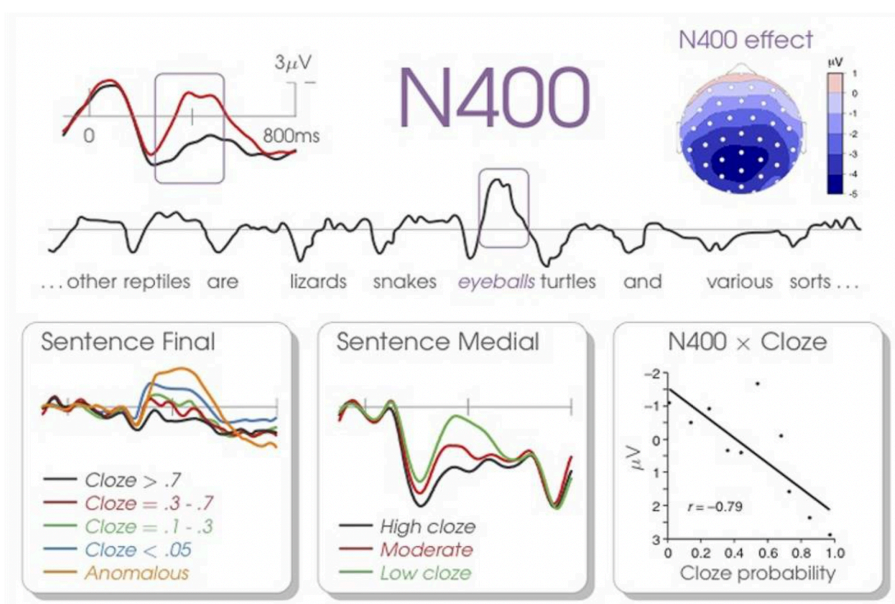

Figure 2. A visual summary of typical N400 effects as a function of word expectancy in a sentence context (i.e., *cloze probability*, measured as the proportion of participants who produce the target word when asked what word they think is most likely to come next; figure taken from a review article on the N400 ERP component (Kutas & Federmeier, 2014))

Example Sentences

High Constraint Sentence (high cloze/expected target word):
The island formed as the result of a volcanic **eruption** thousands of years ago.

Low Constraint Sentence (low cloze/unexpected target word):
The very spoiled child let out a dramatic **eruption** of tears before school.

We used these canonical N400 responses that have been shown to vary based on word expectancy (implemented here by manipulating sentence constraint) as a test case for identifying how language processing varies depending on specific eye movement behaviors. In other words, we isolated this well-documented brain response on different categories of fixation sequences to see if they differed in how much neural language processing is associated with each type of behavioral eye movement decision. These fixation types of interest were (1) single fixation on a word (i.e., cases in which the reader only looked at the word once before moving on), (2) first of multiple fixations (i.e., the first time the reader looked at the word in cases where they decided to then refixate it), and (3) second of multiple fixations (i.e. the refixation itself in case 2; see Figure 3 for a diagram of these different types of fixation behaviors).

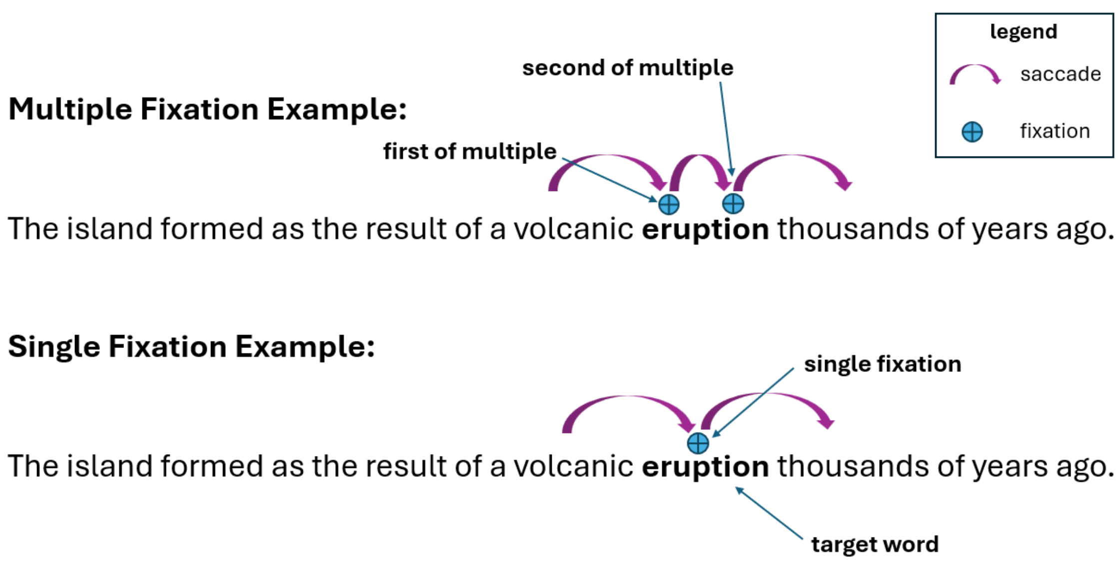

Figure 3. Diagram visualizing the behaviors of interest in the analyzed study and a sample of experimental sentences. *Note:* The target word is presented in boldface in this example, but was presented normally in the experiment.

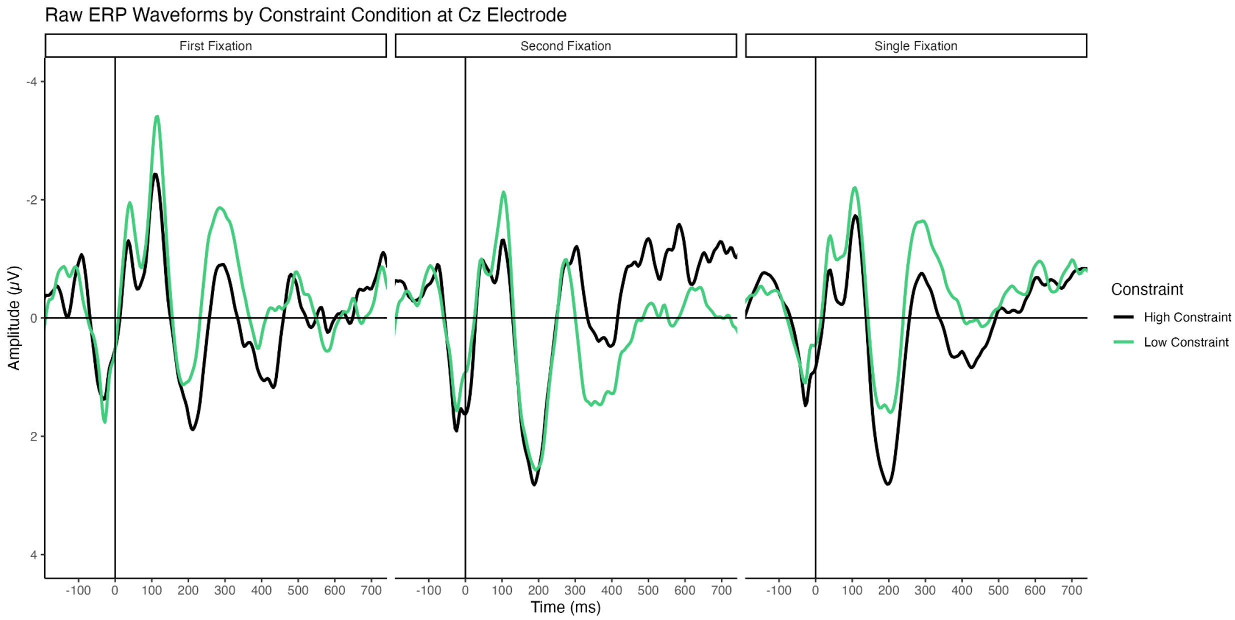

Figure 4. Diagram visualizing the raw results from our initial study on refixations. In each case, the N400 component is the difference between the high constraint and low constraint curves. The second fixation shows a positive deflection of the N400, prompting us to study further.

From our initial study, we found our results for the second fixation to be inconsistent with the existing literature on the N400 component as it showed a positive deflection compared to the other fixations. We reasoned that a possible explanation was the baseline correction of the graphs in the first fixation interfering with the second fixation, as they overlap in time. This influenced the methodology of the current paper: an attempt to see if deconvoluting the first and second fixation (collectively the refixation) would yield more interpretable results.

Hypotheses

Following visual inspection of the unmodeled fixation-related potential (FRP) effects (see Figure 4), we questioned whether the effects of sentence constraint on the second of multiple fixations were being contaminated by effects from the previous fixation, which overlapped temporally with the second fixation. We hypothesized that the observed data pattern may be the result of convolution and that the deconvolution method could potentially correct the otherwise ‘spurious’ effect of the first fixation changing the amplitudes of the N400 on the second fixation.

Our null hypothesis was that there would be no difference between the raw effects and deconvolved effects when taking into account the temporal overlap of the two fixations and associated neural responses; the raw data would represent the true effect of constraint on both the first and second fixation.

Our alternative hypothesis was that correcting for this overlap and for the influence of re-baseline the second fixation response that occurred during the initial first-fixation response using deconvolution would change the overall pattern of the constraint-related N400 response on refixation trials. In particular, we expected that the second fixation response difference between the high and low constraint conditions would be eliminated, or even reversed in polarity, after correcting for convolution from temporal overlap. In the modeled results, this would present as a large shift (towards a more negative voltage N400 response) in the deconvolution modeled *interaction between sentence constraint effect (i.e., high vs. low constraint) and fixation index (i.e., first vs. second of multiple fixations)*. Such a shift in the modeled estimate of the FRP amplitude would indicate that the reversed N400 constraint effect observed in the raw data was largely due to the influence of voltage shifts from the first fixation contaminating the signal rather than the true neural response associated with the second fixation.

Moreover, we also inquired into whether isolating the first and the second of multiple fixations would cause the N400 elicited by the single and first of multiple fixations to remain evenly matched in amplitude (see Figure 4). We also had a secondary alternative hypothesis regarding the influence of the second fixation on the measured response from the first fixation. Because the second fixation often occurred during the measured response from the first, we hypothesized that deconvolution might influence the first fixation effect observed in the raw data. We did not have strong a priori hypotheses about how deconvolution might change the first fixation response, however. We speculated that deconvolution could serve to either shrink or enlarge the N400 constraint effect to the first fixation, so our investigation of changes in this effect was more exploratory. The primary question regarding first fixation effects was whether the magnitude of the N400 remained comparable between the first of multiple fixations and the first single fixations (i.e., when the word was not fixated a second time). Changes in the amplitude based on these fixation types would present as differences between the modelled and unmodelled responses in the *interaction between constraint and fixation type*.

Method

The data used in this analysis was taken from a previous study involving sixty-nine undergraduates at the University of South Florida, recruited either via the Psychology Department's subject pool (and compensated with course credit) or via flyers and mailing lists (and compensated with $16/hr). Participants were between 18 and 35 years of age, native English speakers with normal or corrected-to-normal vision and no history of reading or neurological disorders. We used results from a final sample of 48 participants, after excluding the rest due to participants not meeting the requirements (n = 5), inability to reduce impedance during capping before experiment completion (n = 3), experimenter error in recording data (n = 2), poor eye tracker calibration (n = 1), excessive artifacts and noise in EEG data (n = 7), and less than 15 trials after data processing (n = 3).

Experimental Design and Procedure

In the study, participants were asked to read 120 sentences (60 low-constraint, 60 high-constraint; see example sentences in *Current study rationale and data overview* section above) with either an expected or unexpected target word, and answer comprehension questions intermittently (on ~25% of sentence trials) to ensure they were paying attention. Cloze predictability for our high constraint sentences ranged from 0.4 to 1, verified via sentence norming procedures, where we asked ten participants (not included in the experiment) to complete a set of sentence fragments and calculated the proportion of answers that were our target words (Taylor, 1953). We also ensured the lengths of the target words and their position in the sentence were consistent across trials. We recorded participants' responses to the target word via EEG and eye-tracking.

We recorded right-eye movements using an SR Research Ltd. Eyelink 1000 Plus eye tracking camera in remote desktop mode (sampling rate of 500 Hz). A 3-point calibration at the center of the screen was done at the start of the experiment, to ensure an accuracy within .3° of visual angle at each point. Recalibration was done intermittently if values were greater than or equal to this amount after the drift check on each trial. Participants were seated at a viewing distance of 60 cm from a BENQ XL2540 model LCD monitor with a 240 Hz refresh rate and screen resolution of 1920 x 1080 pixels. EEG was recorded from 30 Ag/AgCl active electrodes (extended 10/20-system) using an actiCAP/actiCHamp electrode cap and amplifier system (Brain Products) with a 500 Hz online sampling rate. No online frequency filters were used during recording.

Data pre-processing

After data collection, data on eye movements were cleaned using the DataViewer program, excluding trials with eye tracking data loss and target word skipping. EEG data was re-referenced offline to the mean of the left and right mastoids, bandpass-filtered between 0.1-50 Hz, with 0.2 - 32.8 Hz half-power (-3dB) cutoffs, using an IIR Butterworth filter. We removed eye movement artifacts using optimized independent components analysis (OPTICAT, version 2020-01-28) as part of the best practices for the field (Dimigen, 2020). For the mass-univariate analysis model, we used two unfold functions (uf_continuousArtifactDetect & uf_continuousArtifactExclude) to find and remove overlapping events from our EEG datasets, setting the threshold at 100μV along our channel of interest (Cz). This was to plot the estimated ERP response without deconvolution (see non-deconvolved graph in Fig. 5). The EEG datasets were time-locked to the start of fixations of the target word between -300ms and 800ms.

We then identified refixations as a sequence of the target word, a saccade, and the target word again in each dataset trial. Single fixations were found as a sequence of the target word, a saccade, and another word in the sentence. We also initialized a fixation index for both fixation types: 'first fixation' and 'next fixation'. For refixations, this simply refers to the initial fixation and the fixation following it. For the single fixation, this refers to the fixation on the target word and the fixation the participant makes on a different word afterward.

Results

Data Analysis

After preprocessing our datasets, we ran the model specified below:

$$FRP \sim 1 + (cat(constraint) \times cat(fix\_type)) + (cat(constraint) \times cat(fix\_index))$$

with binary categorical predictors encoding whether a trial had a high or low-constraint sentence (*constraint: LC, HC*), whether the target word was fixated only once or fixated (*fix_type: refix, single*), and the index of fixations in a refixation (*fix_index: first fix, next fix*). We set LC as the reference level in factor *constraint,* refix as the baseline level in factor *fix_type, and* first_fix as the reference level in factor *fix_next*. Using treatment coding, we can interpret the results as showing the difference from the baseline terms. Below are the terms expressed more explicitly:

- Intercept = the mean FRP amplitude at the levels of high *Constraint*, refixation (*Fixation Type* variable), and first fixation (*Fixation Index* variable)
- constraint = how the FRP differs between high-constraint (HC; intercept) vs. low-constraint (LC) sentences at the intercept levels of *Fixation Type* and *Fixation Index*
- fix_type = how the FRP differs between refixations (intercept) vs single fixations, at the intercept levels of *Constraint* and *Fixation Index*
- (constraint:fix_type) = whether the difference between constraint conditions changes as a function of *Fixation Type* (i.e., refixations (intercept) vs. single fixations), at the intercept level of *Fixation Index*
- (constraint:fix_index) = whether the difference between constraint conditions changes as a function of *Fixation Index* (first (intercept) vs. second of multiple refixations), at the intercept level of *Fixation Type*

By specifying the model this way, we are expressing our modelled deconvolved and non-deconvolved FRPs as the predicted effect for the intercept, the main effects of constraint, fixation type, and fixation index, and the two-way interactions between constraint and the two fixation behavior variables (i.e., Type and Index), explicitly defined as the formula:

$$cat(constraint)\ +\ cat(fix\_type)\ +\ cat(constraint){:}\,cat(fix\_type)$$

and the main and interaction effects between constraint and the two parts of the refixation, also defined similarly. The events were modeled within a time window of -300 to 800 ms, and the waveforms generated were baseline corrected by *subtracting the mean amplitude in the interval from -100 to 0 ms before event onset*. We applied the linear model using deconvolution and mass univariate linear modeling (an approach that fits a linear model over each time point in the response waveform; Dimigen & Ehinger, 2021) to compare differences between deconvolved and non-deconvolved waveforms. We did this by exporting both models' results across our datasets and consolidating them in R for group statistical analyses.

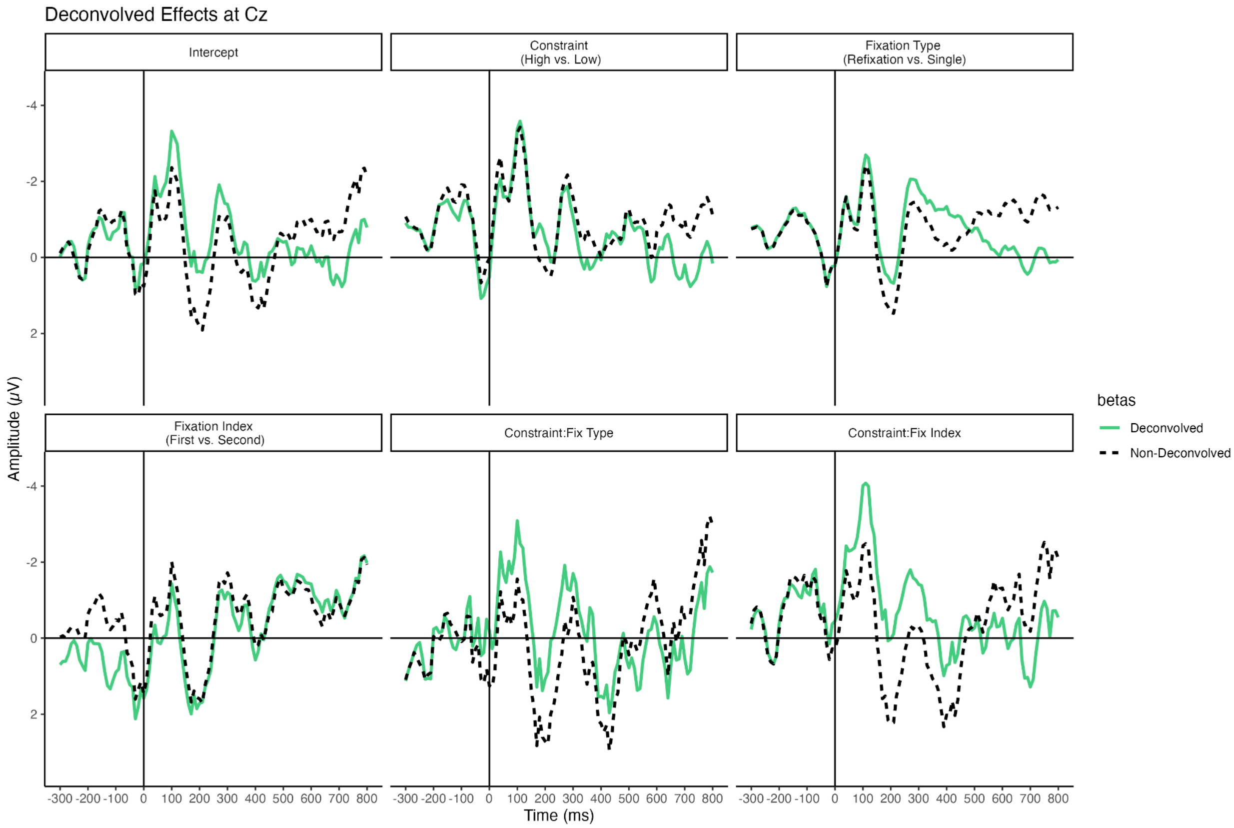

Figure 5. FRP predicted voltage amplitudes at Cz electrode comparing the overall group results when solved by the deconvolution and the mass univariate analysis models.

Each graph in Figure 5 is a sum of the intercept and each modelled effect (with the exception of the top left panel, Intercept, which shows the intercept effect in isolation). The two panels that show the interactions of interest with regard to our hypotheses are the Constraint:Fix Type effect and the Constraint:Fixation Index effect.

Visual inspection of the deconvolved and non-deconvolved effects reveals that the interaction between fixation type and sentence constraint is more negative (indicating a larger N400 constraint effect) for single versus refixations (i.e., the intercept) when the signal is deconvolved. Similarly, the interaction between sentence constraint and fixation index (i.e., first versus second) shows that when the signal is deconvolved, the constraint effect on second fixations is shifted in the negative direction relative to the intercept (i.e., first fixations). These data pattern shifts after deconvolution indicate that the raw data are (1) underestimating the N400 constraint effect of single fixations compared to the first of multiple fixations, and (2) underestimating the N400 constraint effect on second of multiple fixations compared to first of multiple (and even making the effect appear to be reversed in the raw waveforms, which show more negative voltage for high compared to low constraint).

To further visualize these data, we performed a subject-level analysis on a dataset that had a relatively high rate of refixations, so that the chances of overlapping neural responses between fixations are increased, giving the model more data to untangle. We also split the neural activity across fixation types, and created ERP image figures (Delorme & Makeig, 2004). An ERP image figure is a consolidated 2D plot in which every horizontal line indicates neural activity in a trial for a single subject time-fixed relative to the fixations (the zero line). The intensity of the colors represents the strength of the neural responses. We produced two versions of these figures: the raw data and the modeled data, which represents the data excluding subject noise (timings of fixation durations, etc). Modeling these results at the trial level provides a deeper understanding being missed in group averaging, and deconvolution helps us clearly define effects by removing data from the overlapping fixation. For this subject, we observed prolonged N400 responses as fixation durations increase, potentially hinting at refixations being a mechanism to offset shallow word processing.

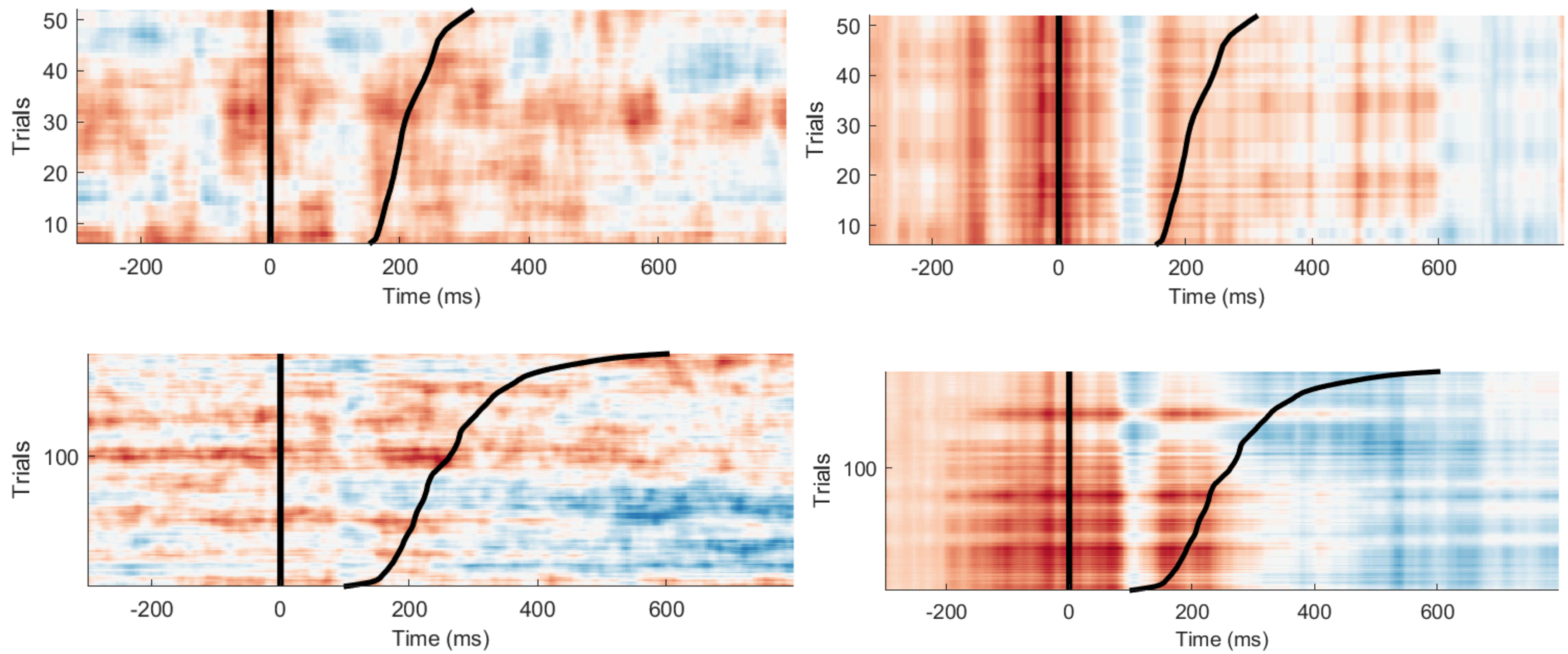

Figure 5. ERP (event-related potential) image figures from a single subject for raw (left panels) and modeled (right panels) data. 56 trials were used to plot the two images at the top, and 178 trials were used to plot the images beneath. *Note:* Trials are aligned to fixation start-end times (the black vertical line) and split across fixation types; the top plotting refixations, and the bottom plotting single fixations. The curved black lines reflect the varying duration times of the fixation type across trials; observable N400 effect spanning longer for longer single fixations than in refixations.

Discussion

Overall, deconvolution has promise in helping us better understand the N400 ERP component alongside other effects that we study. Our results highlight the importance of deconvolution in interpreting effects when overlapping potentials are involved, as the raw data can present unlikely outcomes. As proposed in our hypothesis, we found a polarity change in our ERP waveforms after accounting for first fixation-second fixation overlap, suggesting that the raw data may not have fully factored in the N400 effect posed by the second fixation. This also aligns with previous suggestions stating that overlapping fixations can mask or distort subsequent ERP components (Dimigen & Ehinger, 2021). Moreover, we made a case for a difference between the amplitude of the N400 between a single fixation and the first of multiple fixations by showing a more prolonged N400 effect in the former in a single subject's trial results. However, more group-based analyses would be needed to safely generalize this effect.

Being a relatively new methodology in the field of coregistration, there is still work to be done in consistently interpreting our results and troubleshooting problems that may arise from our model definition itself. In this course of the project, we iterated over many preprocessing scripts and model formulas. The settled-upon method for this paper is what, at the time of writing, the authors reason to align with our understanding of event-related potentials and refixation behaviours, and the least likely to be attributed to any individual idiosyncrasies of the datasets used.

Moreover, we know from the literature that subjects are more likely to make longer fixations during low constraint sentence conditions, so it could be that the longer fixations presented in the first plot of Figure 5 are as a result of this effect. To refine this approach going forward, we will also examine the same ERP plots across constraint conditions to eliminate confounds, and in addition to this, make ERP plots that average across a larger sample space of datasets.

The model ERP plots in Figure 5 also show potential activity of interest before the fixation line. A part of our objectives in our study of refixations is to understand the activity that leads up to it, which would help us conclude whether they are random or executed when conditions are met. We maintain that our report on the current state of this project is sufficient as proof-of-concept that our new methodology provides clearer details on the intricacies of eye movement behavior and word processing, alongside direction on interesting ideas we could probe into in further studies.

References

Degno, F., Loberg, O., Zang, C., Zhang, M., Donnelly, N., & Liversedge, S. P. (2019). Parafoveal previews and lexical frequency in natural reading: Evidence from eye movements and fixation-related potentials.*Journal of Experimental Psychology: General, 148*(3), 453–474. https://doi.org/10.1037/xge0000494

Delorme, A. & Makeig, S. (2004) EEGLAB: an open-source toolbox for analysis of single-trial EEG dynamics, Journal of Neuroscience Methods 134:9-21

Dimigen, O. (2020). Optimizing the ICA-based removal of ocular EEG artifacts from free viewing experiments. NeuroImage, 207, 116117.

Dimigen, O., & Ehinger, B. V. (2021). Regression-based analysis of combined EEG and eye-tracking data: Theory and applications. Journal of Vision 21 (1):3, 1–30, https://doi.org/10.1167/jov.21.1.3.

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(1), 621-647.

Kutas, M., Federmeier, K. D., & Urbach, T. P. (2014). The "negatives" and "positives" of prediction in language. In M. S. Gazzaniga & G. R. Mangun (Eds.), The cognitive neurosciences (5th ed., pp. 649–656). Boston Review.

McConkie, G. W., Kerr, P. W., Reddix, M. D., Zola, D., & Jacobs, A. M. (1989). Eye movement control during reading: II. Frequency of refixating a word. Perception & Psychophysics, 46(3), 245-253.

Rayner, K. (2009). Eye movements and attention in reading, scene perception, and visual search. The Quarterly Journal of Experimental Psychology, 62(8), 1457–1506.

Taylor, W. L. (1953). “Cloze procedure”: A new tool for measuring readability. Journalism quarterly, 30(4), 415-433.