

Modelling Non-linear Relationships in ERP Data Using Mixed-effects Regression with R Examples

Antoine Tremblay

Dalhousie University, Halifax, Nova Scotia, Canada

Aaron J. Newman

Dalhousie University, Halifax, Nova Scotia, Canada

Abstract

In the analysis of psychological and psychophysiological data, the relationship between two variables is often assumed to be a straight line. This may be due to the prevalence of the general linear model in data analysis in these fields, which makes this assumption implicitly. However, there are many problems for which this assumption does not hold. In this paper, we show that in the analysis of event-related potential (ERP) data, the assumption of linearity comes at a cost and may significantly affect the inferences drawn from the data. We demonstrate why the assumption of linearity should be relaxed and how to model nonlinear relationships between ERP amplitudes and predictor variables within the familiar framework of generalized linear models, using restricted cubic splines and mixed-effects regression.

This paper has been written in L^AT_EX using Sweave and R, and the source document is provided as supplementary material. The data, a pdf of this paper, the .Rnw file used to write it, and the R code used to generate all of the analyses, tables, and figures presented here are available in package `EEGLMERdata`. The package is available as supplementary material for this article, and also from <http://hdl.handle.net/10222/22146>. These resources can be used to work through the examples, and potentially act as a starting point for the reader's own forays into mixed-effects analysis.

Keywords: EEG; ERP; nonlinear relationships; LME; restricted cubic splines.

Corresponding author: Antoine Tremblay, NeuroCognitive Imaging Laboratory, Dalhousie University, Halifax, NS B3H 4R2, Canada. Tel.: (902) 494-1911. A.T. was supported during data collection by a Social Sciences and Humanities Research Council of Canada (SSHRC) doctoral fellowship while in the Department of Linguistics at the University of Alberta, Edmonton, Canada (2007–2009). A.T. was supported during analysis and writing by a SSHRC post-doctoral fellowship in the Department of Psychology and Neuroscience at Dalhousie University, Halifax, Canada (2011–2013). A.J.N. was supported by a SSHRC Standard Research Grant.

Introduction

Event-related potential (ERP) datasets are commonly collected in cognitive neuroscience experiments because they offer rich spatio-temporal information about brain activity during perception, cognition, and action. The potential power of ERP data comes with the cost that the datasets tend to be large and quite complex. Thus the richness of the data demands analytical techniques that are appropriate to fully describe both the complexity of the data, and control for factors that might interfere with such analyses. In trying to deal with such complex data, it is very common to assume that the relationship between ERP amplitudes/latencies and an independent variable is linear (i.e., monotonic). For example, in a study that investigates the relationship between age and amplitude of the N1 ERP component, one would assume that a given increase in age is associated with a consistent increase (or decrease) in the amplitude and/or latency of the N1, regardless of whether the increase is from 23 to 24 years old or from 69 to 70 years old.

In reality, there are many problems for which the answer is not a straight line (Rupert, Wand, & Carroll, 2003; Pinheiro & Bates, 2000; Wood, 2006; H. Wu & Zhang, 2006; Hastie & Tibshirani, 1990; Baayen, Kuperman, & Bertram, 2010; Tremblay & Baayen, 2010; Kryuchkova, Tucker, Wurm, & Baayen, 2012) and ERPs are no exception. For example, Carbon, Schweinberger, Kaufmann, and Leder (2005) investigated the effect of “Thatcherized” faces (in which the eyes and mouth regions are turned upside-down) on N170 amplitudes (a negative component occurring between 130 and 200 ms at occipito-temporal scalp sites). Pictures of normal and Thatcherized faces were presented at either 0, 90, or 180° rotation; only when presented upright are these perceived as severely distorted. Carbon et al. (2005) found a nonlinear effect of rotation on N170 amplitudes in response to faces: Upright Thatcherized faces elicited a smaller N170 than Thatcherized faces rotated 90 or 180°. The effect on N170 amplitudes of the latter two orientations, however, did not differ. Boutheina, Coutya, Langer, and Roy (2009) also investigated the effects of picture rotation on N170 amplitudes. They used pictures of normal faces that were rotated 0, 22.5, 45, 67.5, 90, 112.5, 157.5, or 180° and found a relatively complex nonlinear effect on N170 amplitudes and latencies (see panel B of Figure 3 in Boutheina et al., 2009). Numerous other examples of nonlinear effects on ERPs can be found in the literature, in areas such as development (Webb, Long, & Nelson, 2005), rhythm perception (Pablos Martin, Deltenre, Rossion, Hoonhorst, & Colin, 2007), and auditory processing (Inui et al., 2010).

The studies mentioned above did not assume linearity, which is a step forward. However, in keeping with the traditional ANOVA approach to statistical analysis, they tested for nonlinear relationships using the common practice of discretizing the variables of interest. That is, they have reduced inherently continuous measures, such as rotation angle (or stimulus duration, interstimulus interval, magnitude, frequency, etc.) to factor variables with only a few levels. Such a practice is likely attributable to the natural development of cognitive neuroscience from behavioural approaches to studying cognition that traditionally rely heavily on ANOVA models. However, it comes with problems of its own. Indeed, not only does discretization induce bias into the analysis, but it also decreases statistical power and increases the probability of finding spurious associations (Cohen, 1983; McCullum, Zhang, Preacher, & Rucker, 2002). Fortunately, such factorization is neither desirable, nor necessary to detect nonlinear relationships.

Relaxing the assumption of linearity: A set of simple examples

Assuming linearity when it does not hold may lead one (i) to over-look important structure in the data, (ii) to conclude that there is no relationship between two variables when in fact there is one, and (iii) to under-estimate the strength of the relationship between a dependent and an independent variable (Sahai & Ageel, 1997). To illustrate these points, we simulated four data sets (based on the `gamSim` function from R package `mgcv`; Wood, 2012; R Development Core Team, 2013) and compare models that assume linearity to ones that do not (for the sake of this example, we will gloss over how the assumption of linearity was relaxed for the latter analyses). In Figure 1, the panels on the left show the models that were fitted assuming linearity, whereas the ones on the right show the models that were fitted without making this assumption. In the first simulated data set, shown in the top panels (A and B), the dependent (Y) variable does not correlate with the predictor variable, and neither of the two models report that it does. The 2 point difference in AIC indicates that both models are equivalent.¹

Panels C and D show data in which there is a linear relationship between the response and predictor variables, which is captured by both linear and nonlinear models. Indeed, the nonlinear model fit results in a straight line since there is no nonlinearity in the relationship, and the AIC values indicates that both models are just as likely given the data. The simulated relationship between the dependent and independent variables in panels E and F is quadratic. The model for which linearity is assumed (panel E) not only fails to capture the quadratic relationship, but it also falsely concludes that there is no significant relationship between X and Y. The model shown in panel F, however, is able to capture the quadratic trend and comparison of the AIC values for the two models confirms that this is a better fit. In panels G and H, the simulated relationship is more complex, taking on a “wavy” function. Although the model for which linearity is assumed does find that there is a significant relationship between X and Y, the actual nature of the relationship is only part of the underlying truth. This leads the model to under-estimate the strength of the relationship between X and Y, and the accuracy of the prediction varies as a function of X. Conversely, the model for which linearity is not assumed, illustrated in panel H, is able to capture the true nature of the relationship, and correspondingly has a much lower AIC value.

Modelling Non-linear Relationships

Testing for nonlinear relationships can be seen as a natural extension of testing for linear relationships. Linear regression is familiar to virtually anyone with a modicum of training in statistics, and allows one to test for a linear relationship between two variables as shown in the left-hand columns of Figure 1. Thus for example in the analysis of the effects of image rotation in the Carbon et al. (2005) study described above, the levels of rotation could have been treated as values along a continuum of possible rotation angles, rather than as categorical levels. The problem with this approach, as demonstrated in the examples in Figure 1, is that if the relationship is nonlinear then a linear fit will not

¹Smaller AIC values reflect more likely models. By convention, we consider that two models differ if the difference in AIC value is equal to or greater than 5, meaning that the model with the lower AIC is 12 times more likely than the one with the higher AIC.

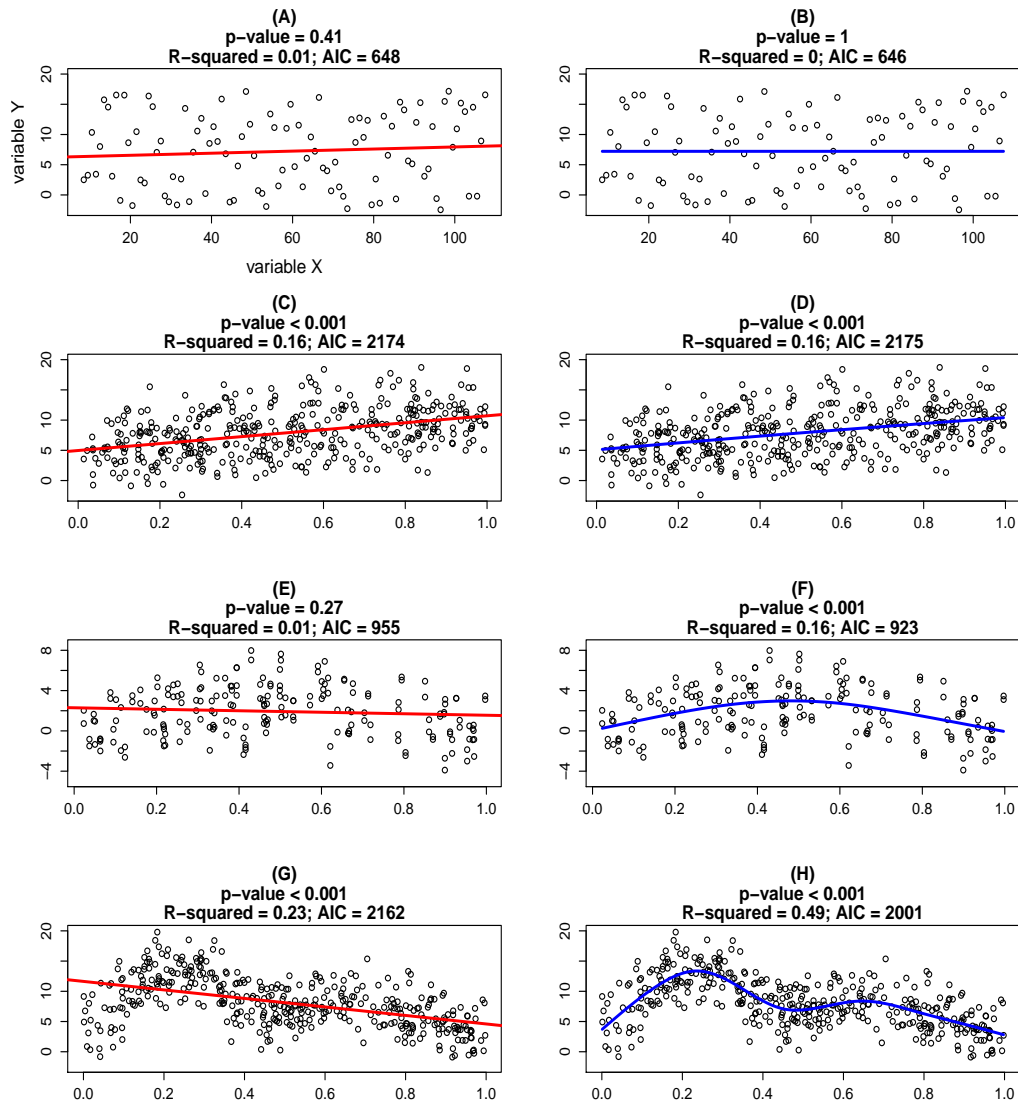


Figure 1. Assuming linearity (left column; red lines) versus not assuming linearity (right column; blue lines).

be optimal. In contrast, pairwise testing of different “levels” of rotation is able to detect which (if any) pairs of levels differ. Certain relationships can also be detected using a set of linear contrast weights, as in Helmert or polynomial contrasts. However, this approach does not generalize easily to many levels of a factor, and because each pairwise or contrast test assumes independence correction for multiple comparisons should be applied — potentially weakening sensitivity to extant effects.

Non-linear relationships can however be modelled without discretization in a number of ways. For example, one can use polynomials, cubic regression splines, thin plate regression splines, Duchon splines, splines on the sphere, P-splines, or Markov random fields. All these

methods essentially apply a series of linear transformations to the predictor variables. In this paper, we use restricted cubic splines (RCS) as implemented by function `rms` from package `rms` (Harrell, 2012) to model nonlinear relationships.² RCS are a set of cubic polynomial functions joined together at points termed “knots”. The RCS function with k knots t_1, \dots, t_k is stated as follows (Harrell, 2001, p. 21, equation 2.26):

$$f(X) = \beta_0 + \beta_1 \mathbf{X} + \beta_2 (\mathbf{X} - t_1)_+^3 + \beta_3 (\mathbf{X} - t_2)_+^3 + \dots + \beta_{(k+1)} (\mathbf{X} - t_k)_+^3 \quad (1)$$

with the conditions that the sections of cubic polynomial must meet at the knots and their slopes must be the same at these points. The basic idea is that the data are represented with a series of cubic polynomials – the $\beta_j (\mathbf{X} - t_k)_+^3$ portions of equation (1) – that are joined together so as to create a continuous curve.

The `rms` function merely serves to transform variables. The actual modelling of the data will be performed with linear mixed-effects regression, which we briefly describe below.

Linear Mixed Effects Regression

Linear mixed-effects regression (LME) is a relatively recent development in statistics, and is a natural tool for modelling repeated measures (H. Wu & Zhang, 2006; L. Wu, 2010). Details about mixed-effects modelling, as well as its potential advantages over rmANOVA, can be found in a number of recent papers and books (e.g., Pinheiro & Bates, 2000; Faraway, 2005; H. Wu & Zhang, 2006; Gelman & Hill, 2007; Baayen, 2008; Baayen, Davidson, & Bates, 2008; Quené & van den Bergh, 2008; Zuur, Ieno, Walker, Saveliev, & Smith, 2009; L. Wu, 2010). LME has been applied to ERP data in several published papers, including Bagiella, Sloan, and Heitjan (2000), Davidson and Indefrey (2007), Moratti, Clementz, Gao, and Keil (2007), Pritchett et al. (2010), Wierda, van Rijn, Taatgen, and Martens (2010), Hsu, Lee, and Marantz (2011), Vossen, van Breukelen, Hermens, van Os, and Lousberg (2011), and Newman, Tremblay, Nichols, Neville, and Ullman (2012).

Like ANOVA, LME also fits within the generalized linear mixed effects model (GLMM). ANOVA and LME differ in that (1) LME can model more complex random-effects structures (e.g., crossed, independent random effects; Baayen et al., 2008; Quené & van den Bergh, 2008); (2) LME uses (restricted) maximum-likelihood methods for parameter estimation (see Pinheiro & Bates, 2000, pp. 62–81 for more details) whereas ANOVA uses the least squares method; (3) LME is robust to violations of sphericity if the correct random effect structure is used (Bagiella et al., 2000; Baayen et al., 2008), eliminating the need to correct for this post hoc using methods that are known to be either overly conservative (Greenhouse-Geisser) or liberal (Huynh-Feldt), and (4) LME enables one to appropriately model imbalanced data (Bagiella et al., 2000; Gelman & Hill, 2007; Baayen et al., 2008), a common situation in ERP data.

To fully make use of these capabilities, data should not be averaged across trials or any other variable. Rather, the LME model should be fitted on the un-averaged (“single trial”)

²We do not use polynomials because “polynomials have some undesirable properties (e.g., undesirable peaks and valleys, and the fit in one region of X can be greatly affected by data in other regions; Magee, 1998) and will not fit adequately many functional forms (Devlin & Weeks, 1986). For example, polynomials do not adequately fit logarithmic functions or ‘threshold’ effects” (Harrell, 2001, p. 18).

data. Beyond allowing robust, accurate estimates of variance, this practice has the added benefit of allowing one to model nonlinear relationships between dependent and independent variables in multi-dimensional space. While at first it may seem odd to describe nonlinear modelling as using “linear” mixed effects, the generalization from the assumption of a linear relationship to a nonlinear one is readily accommodated by LME.

Goals of the Present Study

The goal of this paper is to demonstrate how relaxing the assumption of linearity can lead to better modelling of ERP data. We also demonstrate how to model nonlinear relationships using restricted cubic splines and mixed-effects regression. This paper has been written in \LaTeX using `Sweave` and `R`, and the source document is provided as supplementary material. The data, a pdf of this paper, the `.Rnw` file used to write it, and the `R` code used to generate all of the analyses, tables, and figures presented in this paper are available in package `EEGLMERdata`. The package is available as supplementary material or from <http://hdl.handle.net/10222/22146>. These resources can be used to work through the examples, and potentially act as a starting point for the reader’s own forays into LME analysis. This paper can, however, be read and understood without viewing the accompanying `R` code.

Example ERP Data

The ERP data used here was collected in the context of an immediate free recall experiment described in Tremblay (2009) and Tremblay and Baayen (2010). The goal of this experiment was to investigate whether frequently used four-word sequences, such as *end of the year*, *I don’t really know*, and *at the same time*, may be (de)composed via the application of compositional rules or stored and retrieved as wholes (in which case they should show effects of the frequency of co-occurrence of the four-word sequence). This was achieved by examining whether the frequency of such phrases affected early ERP components such as the anterior N1 (N1a) and the posterior P1. Both of these components peak approximately 110–150 ms after stimulus presentation, and were previously found to be modulated in studies of single words by linguistic variables including frequency and length (Serenó, Rayner, & Posner, 1998; Assadollahi & Pulvermüller, 2003; Hauk & Pulvermüller, 2004; Hauk, Davis, Ford, Pulvermüller, & Marslen-Wilson, 2006; Murphy, Roodenrys, & Fox, 2006; Penolazzi, Hauk, & Pulvermüller, 2007). Tremblay (2009) reported that higher-frequency four-word sequences indeed elicited more negative N1 and less positive P1 amplitudes than low-frequency sequences.

However, additional data were collected in this study that were not analyzed in Tremblay (2009) or Tremblay and Baayen (2010). In the present paper, we investigate whether working memory capacity, the length of the second word of a sequence, and the position in a list where a four-word sequence was presented (i.e., whether it was presented first, second, . . . , fifth, or sixth in a block), and their interaction affected ERPs in the memory task.³

³The second word of a sequence appeared, more often than not, right where the fixation cross was presented (centre of the screen). Thus, the second word of a sequence was the first word that participants saw. It stands to reason that the longer the second word of a sequence, the greater the amplitude of the N1

In the analyses presented here we focused on the N1a as the ERP component of interest. The N1a is known to be sensitive to spatial attention as well as lexical frequency, probability of occurrence, and word length (e.g., Luck, 2005; Murphy et al., 2006; Penolazzi et al., 2007, and references cited therein). For instance, high-frequency words elicit greater N1a amplitudes than low-frequency words, and longer words elicit larger N1a amplitudes than shorter ones. Because in this paper our focus is on demonstrating nonlinear analysis and comparing it to a linear approach, we limited our analyses to a single electrode (Fz, located along the midline, midway between the nose and the vertex of the head) where the N1a component was maximal, and we do not discuss the implications of our results with respect to the cognitive or linguistic literature.

Participants

Ten right-handed female students from the University of Alberta were paid for their participation in the experiment. (Mean age = 23.4; $SD = 1.6$; min/max = 22/27). All were healthy native speakers of English, had normal or corrected-to-normal vision, and did not report any neurological deficits. Participants had a mean handedness score (Oldfield, 1971) of 79.5/100 ($SD = 15.8$). We assessed participants' reading span and working memory capacity (henceforth WMC) using an adaptation of the Daneman and Carpenter (1980) test presented on a PC using E-Prime (Mean WMC score = 73/100; $SD = 10.4$).

Study design

The stimulus list consisted of 432 four-word sequences taken from the *British National Corpus* (Davies, 2004). Some examples are *end of the year*, *I don't really know*, *at the same time*, *I have to say*, *it would be a*, *at the age of*, *this is not a*, *we've got to get*, and *I think it's the*. Frequencies, which were obtained from the *Variations in English Words and Phrases* search engine (Fletcher, 2008), ranged from 0.03 to 105 occurrences per million. The stimuli were divided into 72 blocks. Each block was divided into 18 trials, where in each trial six four-word sequences were randomly presented one at a time in the middle of the screen with an inter-stimulus interval of roughly 4000 ms. Sequences subtended on average $\sim 5^\circ \times 0.4^\circ$ visual angle; the longest four-word string (*becoming increasingly clear that*) subtended $\sim 8^\circ \times 0.4^\circ$ visual angle. At the end of each trial (i.e., after having seen six four-word sequences), participants were prompted to type in as many sequences as they could recall.

EEG Recordings and Processing

Electroencephalogram (EEG) recordings were made with active Ag/AgCl electrodes from 32 locations according to the international 10/20 system (www.biosemi.com/headcap.htm) at the midline (Fz, Cz, Pz, Oz) and left and right hemisphere (Fp1, Fp2, AF3, AF4, F3, F4, F7, F8, FC1, FC2, FC5, FC6, C3, C4, T7, T8, CP1, CP2, CP5, CP6, P3, P4, P7, P8, PO3, PO4, O1, O2), as well as the right and left mastoids, and referenced online at the common mode sense active electrode. Electrodes were mounted on a nylon cap. Eye movements were monitored by electrodes placed above and below the left eye and at the outer canthi of both eyes, which were bipolarized off-line

should be given that longer words elicit larger anterior negativities (Murphy et al., 2006).

to yield vertical and horizontal electro-oculograms (EOG). Analog signals were sampled at 8,192 Hz using a BioSemi (Amsterdam, The Netherlands) Active II digital 24 bits amplification system with an active input range of ± 262 mV per bit and were band-pass filtered between 0.01 and 100 Hz. Note that Biosemi uses active electrodes and therefore can tolerate high scalp impedances.⁴

The digitized EEG was initially processed off-line using **Analyzer** version 1.05 (Brain Products GmbH, Gilching, Germany): It was re-referenced to the average of the right and left mastoids, downsampled to 128 Hz, band-pass filtered from 0.01 to 30 Hz using a forward-backward filter combination where each of the filters was comprised of a two pole zero phase infinite impulse response Butterworth filter, and corrected for eye movements and eye blinks by regressing out the vertical and horizontal EOGs (Gratton, Coles, & Donchin, 1983). The processed signal was then segmented into epochs of 3000 ms (1500 ms before and after stimulus onset). Each epoch was baseline corrected on the 1500 ms segment immediately preceding stimulus onset.

Results

The three predictor variables (fixed effects) under investigation were (1) the position of the four-word sequence in the set of six items presented in a block (Position), (2) the length of the second word in the four-word chunk (Length)⁵, and (3) the working memory capacity of the participant (WMC). These predictor variables were transformed prior to plotting or analysis. We first mean-centered Position, Length, and WMC by subtracting the mean of a variable from each of its values. The main purpose of this centering was to increase numerical stability and remove any spuriously high correlation that may arise between random intercepts and random slopes as well as between fixed and random variables (Hofman & Gavin, 1998; Kreft & De Leeuw, 1998; Baayen, 2008). We first present the results from the behavioural analysis and then move on to the ERP analysis.

Behavioural Results

We investigated whether Length, Position, WMC, and their interactions affected the probability of correctly recalling a four-word sequence. In order to be correctly recalled, a four-word sequence had to be recalled exactly. That is, if the target sequence was *in the middle of*, any response other than *in the middle of* was considered to be incorrect (e.g., *in the middle*, *in the middle and*, *in the middle of a*, or *at the middle of*). However, we accepted minor misspellings such as *in the mdle of* or *n the midle of*.

The optimal generalized LME model (for which linearity was *not* assumed) included main effects of Length ($\chi_{(2)} = 28.2$, $p < 0.001$), Position ($\chi_{(8)} = 512.1$, $p < 0.001$), and WMC ($\chi_{(3)} = 13.3$, $p = 0.004$).⁶ Results of the behavioural analysis are depicted in Figure 2. Length had a small, albeit significant, effect on the probability of correctly recalling a four-word sequence. As illustrated in panel A of Figure 2, the probability of successful recall was

⁴See <http://www.biosemi.nl/forum/viewtopic.php?t=486&highlight=impedance>

⁵The second word was chosen because it most often overlapped the center of the monitor where subjects fixated between trials

⁶These statistics were obtained by way of log-likelihood ratio tests between models with and without the main effects. Moreover, models with the main effects always had an AIC value lower than models without the main effects.

roughly the same when the length of the second word was 1, 2, 3, and 4 letters, but started to decrease from there on as length increased. The Position effect shown in panel B exhibited a large recency effect whereby more recently presented four-word sequences were more likely to be correctly recalled. This possibly reflected an advantage in activation strength for the sequences that were presented more recently within a block (Jones & Oberauer, 2013). Finally, as can be seen in panel C, the greater the working memory capacity of a participant, the more likely she was to correctly remember a four-word sequence.

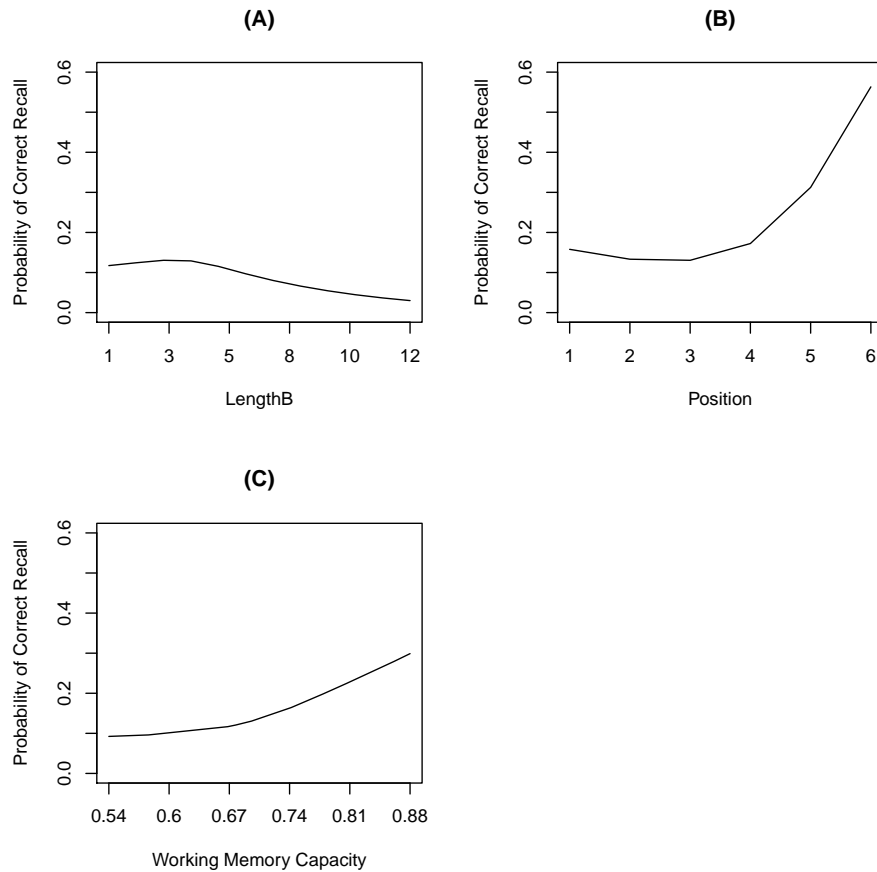


Figure 2. Results of the behavioural analysis. (A): Partial effect of Position on the probability of correctly recalling a four-word sequence. (B): Partial effect of Length. (C): Partial effect of Working Memory Capacity.

ERP Results

An mp4 movie of the scalp topographies through time can be downloaded from <http://hdl.handle.net/10222/22146>. Figure 3 depicts the scalp topography at time $t = 141$ ms where the N1a was maximal. It is apparent from the scalp topography movie, as well as from the bottom panel of Figure 3, that a negative deflection around electrode

Fz — the N1a — began at $t \sim 78$ ms, peaked at time $t = 141$ ms ($\sim -5\mu V$), and returned to baseline at time $t \sim 180$ ms. Our window for analysis was thus chosen as 80–180 ms.

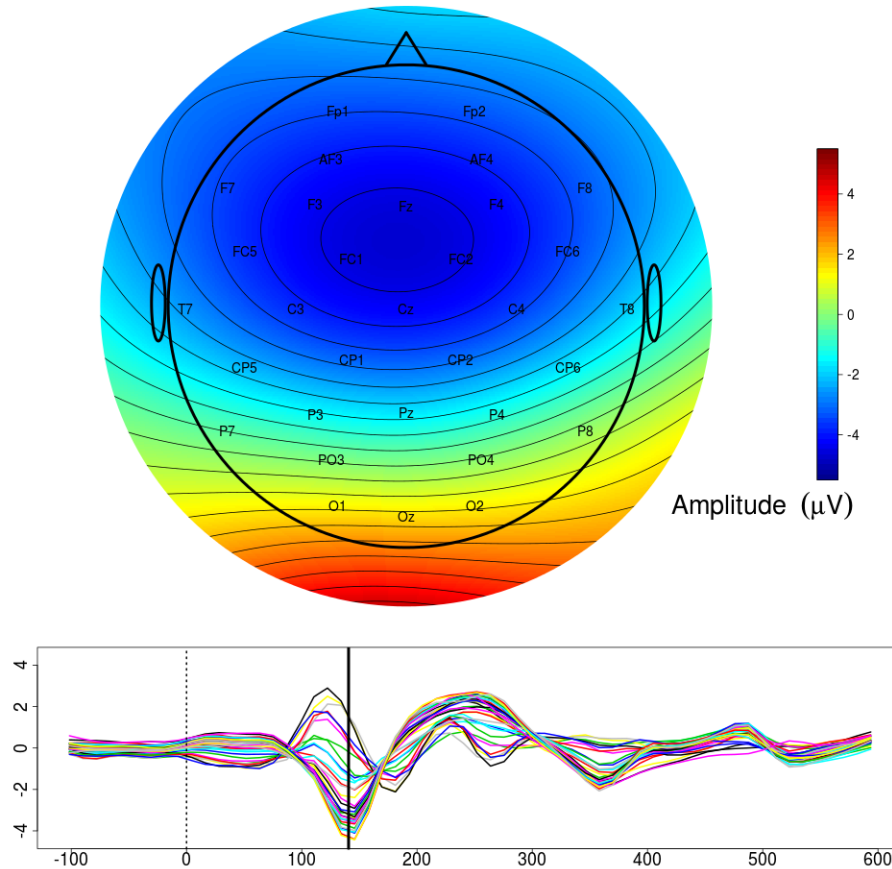


Figure 3. Scalp topography at time $t = 141$ ms where the N1a was maximal. Blue colors indicate negative amplitudes while red colors indicate positive amplitudes. White indicates an amplitude of $0 \mu V$. The bottom panel shows the average waveform of each of the 32 electrodes overlaid on top of each other (each electrodes is graphed with a different color). The x -axis represents time in milliseconds and the y -axis is amplitude in μV . A vertical black bar marks the peak of the N1a component.

Although our main goal in this paper is to demonstrate the application of nonlinear analysis using continuous variables, for descriptive purposes, as well as for comparison with traditional ANOVA-based approaches, we first show the data treated categorically. Thus Position was treated as having 6 levels while Length and WMC were dichotomized as high/low based on a median split. Figure 4 shows waveforms and mean N1a amplitudes for each level of each of these three variables. No effect of Length is apparent in this figure. However, there do appear to be amplitude differences for different levels of Position (middle panel) — in particular, items that appeared first in the lists show a more negative N1a than subsequent positions. As such, there may be a nonlinear relationship between Position

and N1a amplitude. The bottom panel also suggests the possibility of an effect of WMC, where participants with a higher WMC showed a more negative N1a than those with a lower working memory capacity.

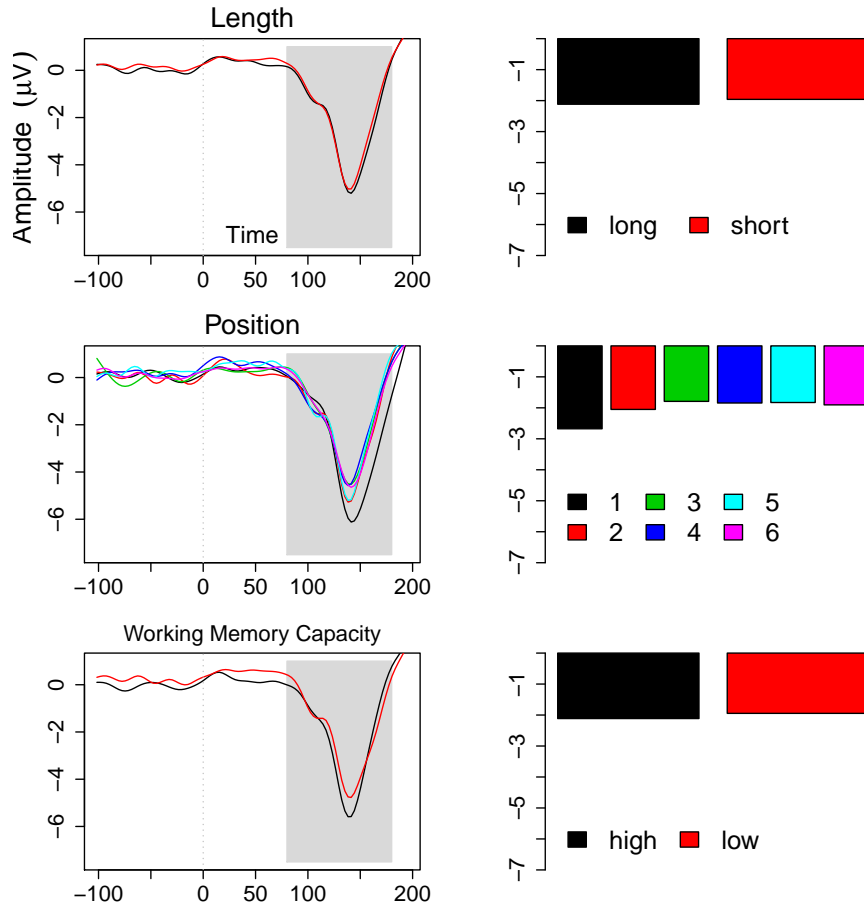


Figure 4. Waveforms (right panels) and mean amplitudes (left panel) of the N1a at electrode Fz, in the 80–180 ms window, plotted separately for discretized levels of each independent variable. The x -axis represents time in milliseconds and the y -axis is amplitude in μV .

Figure 5 shows waveforms and mean amplitudes for two- and three-way combinations of levels of the independent variables (corresponding to the two- and three-way interactions in an ANOVA model). There appears to be a $Length \times Position$ interaction, where long words in first position elicited a more negative N1a than short words in first position. There may be a $Length \times WMC$ interaction and a $WMC \times Position$ interaction. In the former case, shorter words elicited a smaller N1a in participants with a low working memory capacity. In the latter case, the Position effect on N1a amplitudes was stronger in participants with high than low working memory capacity. Finally, there is possibly a three-way interaction (bottom panel), where long words in the first position elicited a more negative N1a in participants with low than high working memory capacity (the leftmost black and red bars).

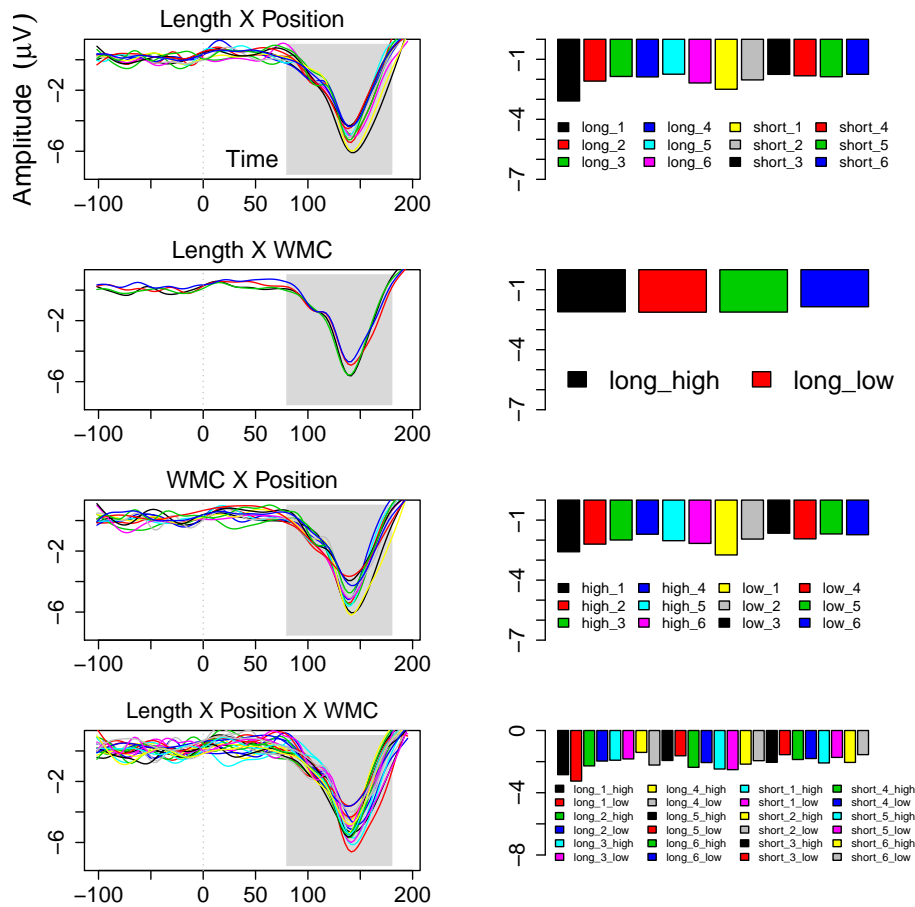


Figure 5. Two-way and three-way interactions between dichotomized levels of *Length*, *Position* and *WMC*. Details are as in Figure 4.

LME Analysis Assuming Linearity

Our first LME analysis simply treated each variable as continuous. Results are shown in Table 1. The model assuming linearity suggests that there was only a significant main effect of Position. While it is notable that a main effect for Position was found under the assumption of linearity, Figure 4 suggested that this variable had a nonlinear relationship with N1a amplitude. Describing the source of this main effect would necessitate drawing inferences over the six levels of Position from a relatively large number of pairwise comparisons.⁷

Fitting a single nonlinear function to Position could simplify the description and interpretation of the effect without recourse to cumbersome post hoc testing. In addition, it may be the case that the assumption of linearity missed important structure in the data. In the next section, we explore LME models for which linearity is *not* assumed.

⁷While contrasts such as Helmert could also describe this relationship somewhat more parsimoniously, we had no a priori assumption as to the shape of the relationship between Position and N1a amplitude so the choice of appropriate contrast weights would have to be made post hoc.

Table 1

Electrode Fz – 80 to 180 ms. Probability (p) values from model for which linearity is assumed. Pos = Position. Lngth = Length, WMC = Working Memory Capacity.

Pos	Lngth	WMC	Pos:Lngth	Pos:WMC	Lngth:WMC	Pos:Lngth:WMC
< .001	0.244	0.713	0.602	0.139	0.105	0.305

Moving Away from the Assumption of Linearity

The possibility of nonlinear relationships between our predictor variables and N1a amplitude is supported by the Figure 6, in which mean N1a amplitude at each sampled level of each variable are plotted, along with a form of best-fitting nonlinear functions (lowess smooths).⁸ The Length effect (top left panel) appears to be curved with relatively constant (and small) N1a amplitudes from values 1 to roughly 6, after which N1a amplitude increases with increasing word length.⁹ The effect of Position (bottom left panel) shows an opposite pattern. The amplitude of the N1a is most negative in the first position, less so in second position, and then flattens out in the third through sixth positions – even becoming slightly more negative at the last position. The WMC effect (top right panel) is more complex, with an initial increase in N1a amplitude from the lowest to low-middle WMC levels, followed by a decrease and then a subsequent increase for people with the highest WMC.

As introduced above, in our LME analysis we modelled these nonlinear relationships with restricted cubic splines (implemented by function `rCs` in R). As a first step these functions needed to be parameterized. That is, we needed to determine how many knots we would use for each predictor. For each cubic spline, we used 2 knots plus the number of inflection points found in the lowess curves (i.e., the red lines in Figure 6). Therefore, for Length and Position we used a restricted cubic spline function of degree 3 (i.e., three knots, one at the 1st, 50th, and 100th quantiles of each variable’s distribution), and for WMC a degree of 4 (i.e., knots at the 1st, 25th, 75th, and 100th quantiles). Setting the knots at quantile locations ensured that each “window” contained an equal number of data points (this is automatically done in function `rCs`). The choice of the number of knots was driven by the shape of the empirically-derived lowess fits (i.e., not driven by investigator assumptions), with additional knots allowing for a more wiggly fit, if necessary. In cases where this degree of fit was unwarranted, the parameters fit to the functions simply allow for a flatter (less convoluted) function.

The base model had the form *Amplitude as a function of* $rCs(Position, 3) \times rCs(Length, 3) \times rCs(WMC, 4)$, that is, a three way interaction between restricted cubic splines for Position, Length, and WMC. This model was much more likely than the model for which linearity was assumed (model not assuming linearity: $df = 51$, $AIC = 355859$; model assuming linearity: $df = 13$, $AIC = 356270$; $\chi^2 = 487.3$, $df = 38$, $p < 0.001$). Results of this model appears in Table 2. The plot of the three-way interaction is shown in Figure 7.

⁸A lowess smooth uses locally-weighted polynomial regression. ‘Local’ is defined as a window of a certain span around each data point (e.g., 50 data points to the left and to the right). Each data point within a window is influenced (i.e., weighted) by its “neighbours”. See `?lowess` in R for more details.

⁹Because the N1a has a negative polarity, we refer to more negative amplitudes as “increases”.

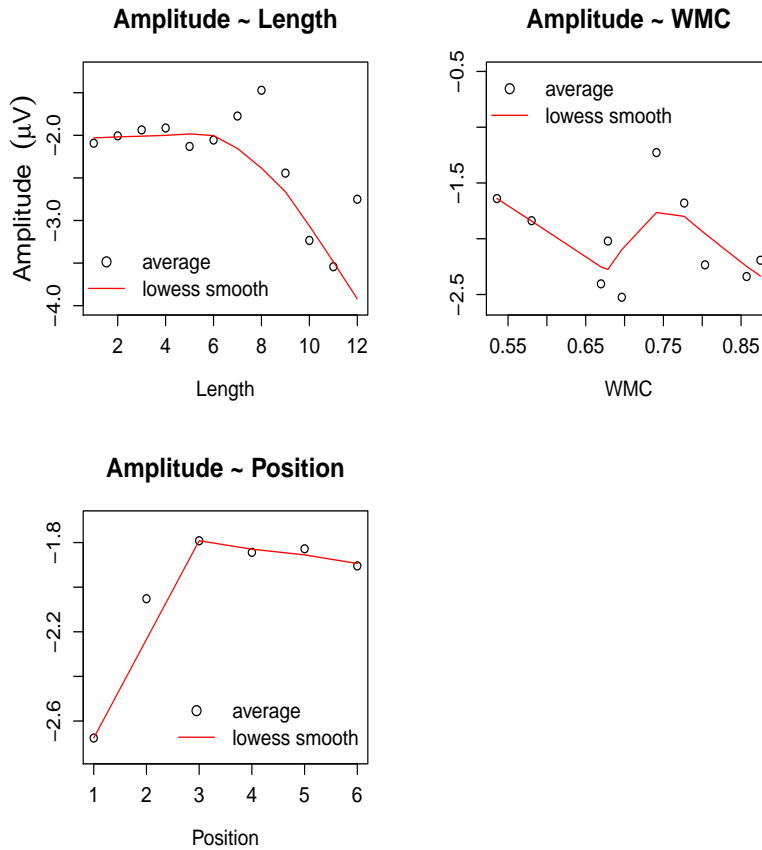


Figure 6. N1a amplitude averages as a function of each one of the predictor variables, treated as continuous (in contrast with the discretized forms of these variables shown in Figure 4). The black circles are the mean amplitude at each measured level of each variable. The red lines are `lowess` smooths of the averages.

Table 2

Probability (p -)values from the model for which linearity was not assumed. *Pos* = Position. *Lngh* = Length, *WMC* = Working Memory Capacity. *rcs* = restricted cubic spline.

Pos	Lngh	WMC	Pos:Lngh	Pos:WMC	Lngh:WMC	Pos:Lngh:WMC
< .001	0.694	0.893	0.182	< 0.001	0.3	0.007

As in the analysis assuming linearity, here we found a significant main effect of Position. This confirmed the pattern observed in both Figures 4 (discretized variables) and 6 (continuous variables) whereby N1a amplitudes were largest for words in the first position but were more or less consistent across subsequent positions. The lack of significant main effects for Length or WMC are also consistent with the linear analysis, suggesting that although the nonlinear curves shown in Figure 6 appeared to describe the data more closely than a straight line would have, nevertheless these relationships were not statistically

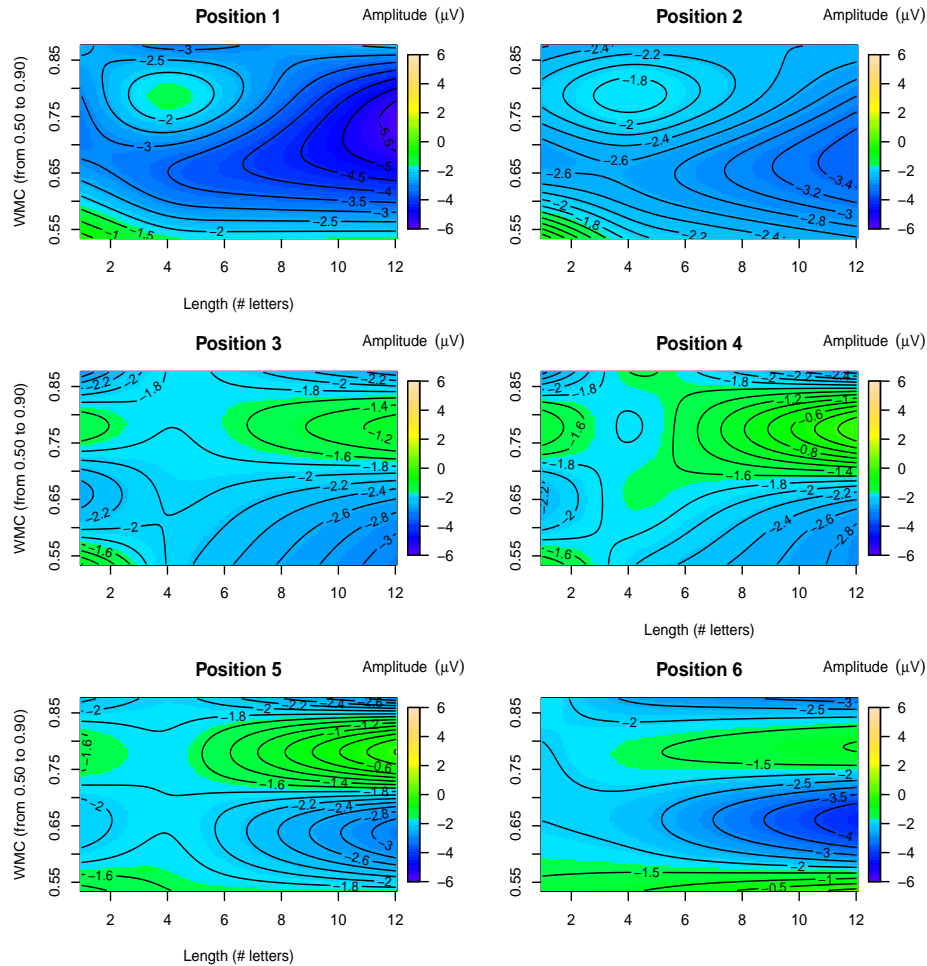


Figure 7. *Length* \times *WMC* interaction of the best-fitting model (*m6*) from the nonlinear analysis. Each panel shows the *Length* \times *WMC* interaction at one of the six levels of Position. The amplitude of the N1a is shown using the same color coding as in Figure 3, with the scale shown in the figure. The small numbers on the black lines are isovoltage lines with the voltage in microvolts provided.

reliable.

More striking differences between the linear and nonlinear analyses reside in the interaction structure of the models. The linear model did not include any significant interactions, while in the nonlinear model both the two-way *Position* \times *WMC* and the three-way *Position* \times *Length* \times *WMC* interactions were significant. Recall that these interactions were predicted above based on our examination of the plots shown in Figure 5. These interactions are provided in Figure 7. To display the three-way interaction, we have plotted the *Length* \times *WMC* interaction at each level of Position. Given the presence of a three-way interaction, we did not attempt to interpret the two-way *Position* \times *WMC* interaction since this is necessarily conditioned on the further interaction of these variables by Length.

Figure 7 shows a complex relationship between these three variables. The stronger

N1a at Position 1 is clear from the larger area of dark blue in the panel for this position, however the interaction shows that this effect was maximal for the longest words, and further was greater for people with WMC in the middle of the sampled range, with somewhat smaller N1a amplitudes at both higher and lower WMC (though because fewer participants had extreme values on any given measure, parameter estimates are less robust at both ends of the WMC range). The nonlinear relationship between Length and N1a amplitude observed in Figure 6 is evident in these plots as well – the spacing between the isovoltage lines is greater at shorter lengths, indicating a slower rate of amplitude increase, compared to higher values of Length. A similar relationship between the three variables is evident in the panel for Position 2, though with lower amplitudes overall. At later positions, however, a different relationship between WMC and N1a amplitude emerged. While N1a amplitude still increased in a nonlinear fashion with Length, at Positions 3–6 people with lower WMC showed larger N1a amplitudes than those with higher WMC. Indeed, at least up to Position 4 people with the lowest WMC showed consistent N1a amplitudes ($\sim 3 \mu\text{V}$ for longest words), while at later levels of position people with higher WMC showed little evidence for an N1a at all (voltage close to $0 \mu\text{V}$).

To summarize these results, we can say that the N1a was robust across subjects at early positions in each sequence of four-word chunks, and larger for longer words. However, as more chunks were presented within a block, N1a amplitudes for people with higher WMC decreased to near zero, while the N1a persisted in people with lower WMC through all six chunks within a block.¹⁰ These results are consistent with the task demands of the study and what we know about working memory and attention. As noted earlier, the N1a is associated with attention, as attended items typically elicit higher N1a amplitudes (e.g., Luck, 2005; Penolazzi et al., 2007, and references cited therein). In this study, participants were required to recall as many of the four-word chunks as possible at the end of each block of six chunks. If we take the N1a as indexing the amount of attentional resources allocated to the task of storing each four-word sequence into short-term memory for later recall, then we can interpret the effects of length as indicating that more attentional resources were required to encode and retain longer words in short-term memory. The effect of Position is consistent with the “primacy” effect in working memory (Deese & Kaufman, 1957) whereby the first-presented items in a list receive more attention and tend to be better-remembered. Likewise, the effects of WMC indicate that people with lower working memory capacity needed to allocate more attentional resources to encoding chunks presented later within a block, particularly for longer words. People with higher WMC capacity, on the other hand, actually appeared to allocate less attentional resources to chunks in later positions, especially those with longer words.

In the behavioural analysis, we had found that Position, LengthB, and WMC affected the probability of a four-word sequence being correctly recalled. More specifically, (i) more recently presented sequences were more likely to be correctly recalled, (ii) sequences for which the second word was longer were less likely to be correctly recalled, and (iii) participants with a higher working memory capacity were more likely to correctly recall a sequence. Our ability to relate the ERP results back to behaviour, however, is seriously compromised

¹⁰Although at Position 6, voltage values at the lowest WMC decreased relative to slightly higher WMC values, as noted earlier the pattern at the tails of WMC must be viewed with skepticism as only one participant had the lowest WMC score.

by the fact that we conflated successfully- and unsuccessfully-recalled sequences. In the following section, we investigate whether the recall status of a sequence affected N1a amplitudes and whether correctly and incorrectly recalled sequences were differentially affected by Position, Length, and WMC.

Interactions with a Dichotomous (Factor) Variable

The preceding section demonstrated a nonlinear analysis of three continuous predictor variables, and how a three-way interaction can be visualized and interpreted. In some cases, however, some predictors are naturally factorial rather than continuous. For example, in the present study participants performed a later recognition judgement task — so accuracy on this task for each chunk could be included as an additional, factorial (correct/incorrect) predictor. Inclusion of factorial predictors is readily handled within the LME framework, and using the *lmer* function this is achieved simply by ensuring that the factorial variable is treated as type “factor” (as opposed to “numeric” as continuous variables are). The additional inclusion of the factorial variable Recalled allowed us to address two additional questions: (1) Is there a relationship between N1a amplitudes and the successful (or unsuccessful) recall of a four-word sequence, and (2) in the event where such a relationship exists, is it modulated by a participant’s working memory capacity, the length of the second word of a sequence, and/or the position where a four-word sequence was presented?

The best fitting model had the form *Amplitude as a function of rcs(Position, 3) × rcs(Length, 3) × rcs(WMC, 4) × Recalled*. It was much more likely than a model with the same fixed- and random-effect structure, but for which linearity was assumed (model not assuming linearity: AIC = 355926, $df = 87$; model assuming linearity: AIC = 356274, $df = 22$; $\chi^2 = 478.4$, $df = 65$, $p < 0.001$). The four-way interaction was significant ($F(12, 50970) = 2.3$, $p = 0.0061$), thus answering “yes” to both of the questions posed above. To visualize this four-way interaction, and considering that the fourth factor, Recalled, was dichotomous, we can plot three-way interaction plots as we did for the previous model, but separately for correctly- and incorrectly-recalled chunks. These appear as Figures 8 and 9, respectively. As a means of data reduction, we have also plotted the difference between these two maps in Figure 10.

Keeping in mind that the plots of the data separated by Recalled are each based on less data than those collapsed across levels of this variable, and are thus likely less robust (again, especially at extreme values of WMC), we can see similar patterns as in Figure 7, such as nonlinearly increasing amplitude with Length and at higher levels of WMC particularly at later Positions. However, overall we can observe more negative amplitudes for recalled chunks in Figure 8 than for non-recalled chunks in Figure 9 — in particular for people with lower levels of WMC and at lower lengths. This is largely consistent with the behavioural results, where it was found that four-word sequences for which the second word was shorter were more often successfully recalled and that participant with a higher working memory capacity had a greater likelihood of correctly recalling a sequence. Additionally, while a primacy effect was noted in the previous analysis, when we look exclusively at correctly-recalled items or at the difference plot in Figure 10 we can see a pattern that seems to correspond to the “recency” effect as well, whereby N1a amplitudes were particularly large for correctly-recalled items at Position 6, again showing the nonlinear increase with Length. Note that this recency effect is consistent with the one found in the behavioural analysis.

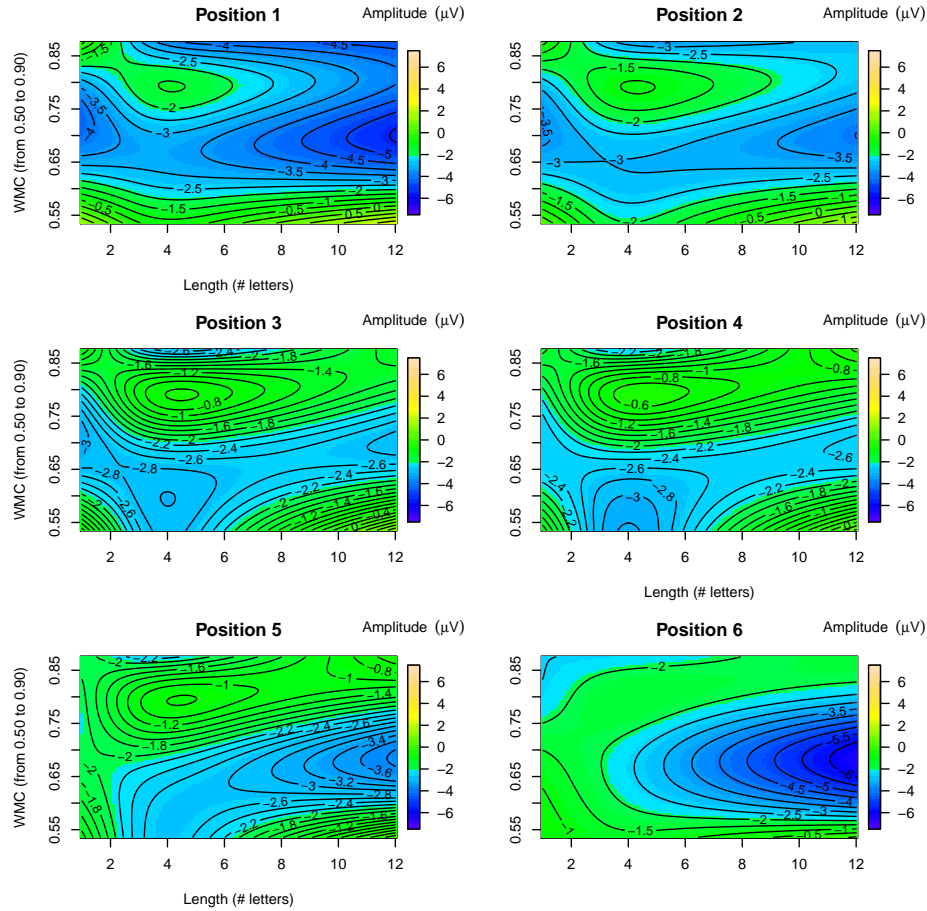


Figure 8. The three way interaction $Position \times Length \times WMC$ for correctly recalled sequences. Details are as for Figure 7.

Overall, these data suggest that while attentional resources (as indexed by the N1a) are not a necessary requirement for successful recall, these resources seem to be allocated preferentially to more difficult items and to a greater degree in people with more limited WMC, and such allocation correlates with later successful recall.

Discussion

Relaxing the assumption of linearity enabled us to gain a better understanding of the data by capturing important structure that was overlooked in the model for which linearity was assumed. Indeed, only a main effect of Position was uncovered in this model whereas the one for which linearity was not assumed revealed a significant main effect of Position in addition to significant $Position \times WMC$ and $Position \times Length \times WMC$ interactions. Moreover, even if these interactions would have reached significance in the model assuming linearity, the association strength between N1a amplitudes and these interactions (as indexed by the percentage of deviance explained) would have been much lower than in the one for which linearity was not assumed, as can be seen in Table 3. While the $Position \times$

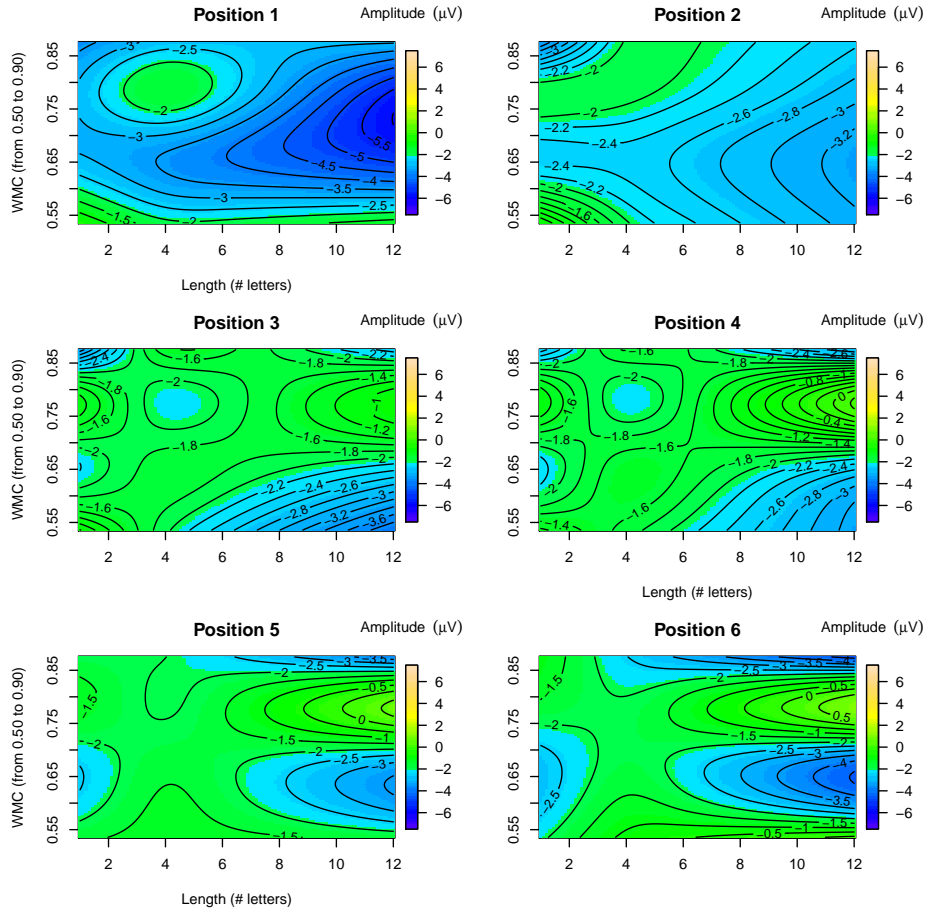


Figure 9. The three way interaction $Position \times Length \times WMC$ for incorrectly recalled sequences. Details are as for Figure 7.

WMC interaction in the model for which linearity was not assumed accounted for 4 times more deviance than in the model for which linearity was assumed, the $Position \times Length \times WMC$ in the former model accounted for 26 times more deviance than same interaction in the latter model.

Finally, we were able to push our understanding of the data even further by allowing the $Length \times WMC$ smooth surfaces at each Position to vary according to whether a four-word sequence was correctly or incorrectly recalled. This four-way interaction was significant even after accounting for subject- and item-wise random intercepts and restricted cubic splines for Position and WMC.

In spite of the clear potential advantage of nonlinear analyses in capturing the true structure of the data, it is important to recognize limitations of this technique. Firstly, the relationship between two variables may in fact be best described by a straight line. In this case, nonlinear analysis is unnecessary; this will be recognized however if model comparison testing is used to determine whether nonlinear fitting is warranted. Secondly, with complexity of analytical model comes complexity in interpretation. In the present case

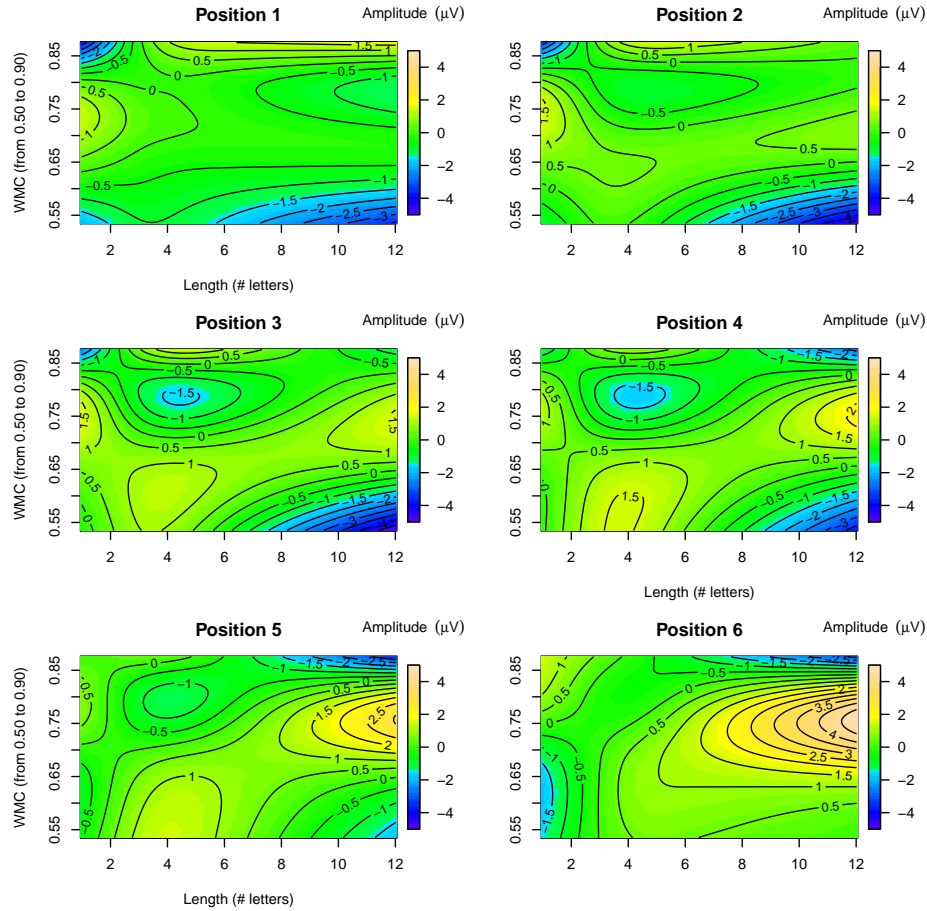


Figure 10. The four way interaction $Position \times Length \times WMC \times Recalled$ – incorrectly minus correctly recalled. Details are as for Figure 7.

we have demonstrated that this complexity is not necessarily uninterpretable however, and indeed the results provide a much richer understanding of the data than could be provided by linear assumptions. A third issue, that we have not detailed thus far, concerns what sort of linear model to fit and what assumptions to make in this regard.

A crucial step in modelling nonlinear relationships between two variables is parametrizing the splines — determining how many knots they will have. The approach we used here was mainly based on visual inspection of averages and lowess smooths. A drawback with this procedure is that it is essentially arbitrary and open to investigator bias. A more data-driven method to determine the correct parametrization of the splines would be to fit a series of successively more complex models (i.e., with different numbers of knots) and select the one with the lowest AIC. If there is only one predictor variable, this may be done relatively easily. However, the space of possible models increases drastically as the number of predictors in the model increases. For instance, if we set the upper bound on the number of knots to 5 for each of three predictors, we would have to fit 125 different models; for a maximum of 10 knots we would require comparisons among 1000 model fits.

Table 3

Percentage of the deviance explained by each term of the model assuming linearity and the one not assuming linearity. The interactions in bold face correspond to ones that were significant in the model for which linearity was not assumed.

	Assuming Linearity	Not Assuming Linearity
Position	0.0255%	0.0211%
Length	0.0024%	0.0013%
WMC	0.0002%	0.0011%
Position \times Length	0.0005%	0.0108%
Position \times WMC	0.0038%	0.0154%
Length \times WMC	0.0046%	0.0125%
Position \times Length \times WMC	0.0018%	0.0473%

An alternative, data-driven approach is to avoid the necessity of performing such a search by using semi- or non-parametric methods such as generalized additive modelling (GAM) and generalized additive mixed-effects modelling (GAMM). The GA(M)M algorithm as implemented in package `mgcv` (Wood, 2012) strikes a balance between under- and over-smoothing the data by using a fixed number of knots and using penalized regression where the least squares objective is augmented by a “wiggleness” penalty. The value of this wiggleness penalty is determined, for example, by way of generalized cross validation (GCV). Crucially, the user only needs to specify an upper-bound on the number of knots to be used and GA(M)M determines how many knots it actually needs to model the data. That is, the exact number of knots passed to the cubic regression spline function is not generally critical in this framework. The only concern is to provide a large enough upper-bound so as to enable GA(M)M to represent the underlying “truth” reasonably well. Although we do not explore the modelling of ERP data using GA(M)M here, an example of how it performs was provided in Figure 1. Indeed, each one of the models for which linearity was not assumed (the right column) were fitted using GAM. The upper-bound on the number of knots was set to 10. In panel B, GAM determined that the number of knots to be used was equal to 0. In panel D, it settled on a linear relationship between X and Y. While it deemed that 3 knots were necessary to model the simulated quadratic curve in panel F, it determined that 7 knots were required to capture the nonlinear relationship in panel H. We refer the interested reader to Wood (2006), Faraway (2005), Keele (2008), and Zuur et al. (2009) for more details on generalized additive (mixed-effects) modelling.

Conclusion

In order to deal with the complexity of ERP data, it is very common to assume that the relationship between ERP amplitudes/latencies and independent variables is linear. This assumption does not always hold, however. Although in some ERP studies the assumption of linearity has been relaxed, these studies have typically employed the common practice of discretizing continuous variables, which has been shown elsewhere to be suboptimal.

In this paper, we have demonstrated how nonlinearities in ERP data can be modeled using restricted cubic splines and mixed-effects regression in R. This revealed that the

assumption of linearity may lead one to draw incorrect inferences from his/her ERP data. More specifically, we have illustrated how the assumption of linearity may (i) increase the probability of false negatives, (ii) if a relationship is obtained, provide a limited or incorrect representation of the true relationship between variables, and (iii) under-estimate the strength of the association between variables.

References

- Assadollahi, R., & Pulvermüller, F. (2003). Early influences of word length and frequency: a group study using MEG. *NeuroReport*, *14*, 1183–1187. doi: 10.1097/01.wnr.0000075305.76650.60
- Baayen, R. H. (2008). *Analyzing linguistic data: A practical introduction to statistics using R*. Cambridge, U.K.: Cambridge University Press.
- Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *J Mem Lang*, *59*, 390–412. doi: 10.1016/j.jml.2007.12.005
- Baayen, R. H., Kuperman, V., & Bertram, R. (2010). Frequency effects in compound processing. In S. Scalise & I. Vogel (Eds.), *Compounding*. (pp. 257–270). Amsterdam and Philadelphia: John Benjamins.
- Bagiella, E., Sloan, R., & Heitjan, D. (2000). Mixed-effects models in psychophysiology. *Psychophysiology*, *37*, 13–20. doi: 10.1017/S0048577200980648
- Boutheina, J., Coutya, J., Langer, C., & Roy, S. (2009). From upright to upside-down presentation: A spatio-temporal ERP study of the parametric effect of rotation on face and house processing. *BMC Neurosci*, *10*, 1–17. doi: 10.1186/1471-2202-10-100
- Carbon, C.-C., Schweinberger, S. R., Kaufmann, J. M., & Leder, H. (2005). The Thatcher illusion seen by the brain: An event-related brain potentials study. *Cognitive Brain Res*, *24*, 544–555. doi: 10.1016/j.cogbrainres.2005.03.008
- Cohen, J. (1983). The cost of dichotomization. *Appl Psych Meas*, *7*, 249–254. doi: 10.1177/014662168300700301
- Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *J Verb Learn Verb Be*, *19*, 450–466. doi: 10.1016/S0022-5371(80)90312-6
- Davidson, D., & Indefrey, P. (2007). An inverse relation between event-related and time-frequency violation responses in sentence processing. *Brain Res*, *1158*, 81–92. doi: 10.1016/j.brainres.2007.04.082
- Davies, M. (2004). *BYU-BNC: The British National Corpus*. Available online at <http://corpus.byu.edu/bnc>.
- Deese, J., & Kaufman, R. (1957). Serial effects in recall of unorganized and sequentially organized verbal material. *J Exp Psychol*, *54*, 180–187. doi: 10.1037/h0040536
- Devlin, T., & Weeks, B. (1986). Spline functions for logistic regression modeling. In I. SAS Institute (Ed.), *Proceedings of the Eleventh Annual SAS Users Group International Conference*. (pp. 646–651). Cary, NY.
- Faraway, J. J. (2005). *Extending the Linear Model with R: Generalized Linear, Mixed Effects and Nonparametric Regression Models*. New York: Chapman & Hall.
- Fletcher, W. H. (2008). *Phrases in english*. Available online at <http://pie.usna.edu/index.html>.
- Gelman, A., & Hill, J. (2007). *Data analysis using regression and multilevel/hierarchical models*. Cambridge: Cambridge University Press.
- Gratton, G., Coles, M., & Donchin, E. (1983). A new method for off-line removal of ocular artifact. *Electroenr Clin Neuro*, *55*(4), 468–84. doi: 10.1016/0013-4694(83)90135-9
- Harrell, F. (2001). *Regression modeling strategies*. Berlin: Springer.
- Harrell, F. (2012). rms: Regression modeling strategies [Computer software manual]. Retrieved from <http://CRAN.R-project.org/package=rms> (R package version 3.5-0)
- Hastie, T., & Tibshirani, R. (1990). *Generalized Additive Model Regression*. New York: Chapman & Hall.

- Hauk, O., Davis, M., Ford, M., Pulvermüller, F., & Marslen-Wilson, W. (2006). The time course of visual word recognition as revealed by linear regression analysis of ERP data. *NeuroImage*, *30*, 1383–1400. doi: 10.1016/j.neuroimage.2005.11.048
- Hauk, O., & Pulvermüller, F. (2004). Effects of word length and frequency on the human event-related potential. *Clin Neurophysiol*, *115*, 1090–1103. doi: 10.1016/j.clinph.2003.12.020
- Hofman, D. A., & Gavin, M. B. (1998). Centering decisions in hierarchical linear models: Implications for research in organizations. *J Manage*, *24*, 623–641. doi: 10.1016/S0149-2063(99)80077-4
- Hsu, C.-H., Lee, C.-Y., & Marantz, A. (2011). Effects of visual complexity and sublexical information in the occipitotemporal cortex in the reading of Chinese phonograms: A single-trial analysis with MEG. *Brain Lang*, *117*, 1-11. doi: 10.1016/j.bandl.2010.10.002
- Inui, K., Urakawa, T., Yamashiro, K., Otsuru, N., Nishihara, M., Takeshima, Y., ... Kakigi, R. (2010). Non-linear laws of echoic memory and auditory change detection in humans. *BMC Neurosci*, *11*, 1–13. doi: 10.1186/1471-2202-11-80
- Jones, T., & Oberauer, K. (2013). Serial-position effects for items and relations in short-term memory. *Memory*, *21*, 347–365. doi: 10.1080/09658211.2012.726629
- Keele, L. (2008). *Semiparametric Regression for the Social Sciences*. New York: Chapman & Hall/CRC.
- Kreft, I., & De Leeuw, J. (1998). *Introducing Multilevel Modeling*. London: Sage.
- Kryuchkova, T., Tucker, B. V., Wurm, L. H., & Baayen, R. H. (2012). Danger and usefulness are detected early in auditory lexical processing: Evidence from electroencephalography. *Brain Lang*, *122*(2), 81–91. doi: 10.1016/j.bandl.2012.05.005
- Luck, S. J. (2005). *An introduction to the event-related potential technique*. Cambridge, MA: MIT Press.
- Magee, L. (1998). Nonlocal behavior in polynomial regressions. *Am Stat*, *52*, 20–22. doi: 10.2307/2685560
- McCallum, R. C., Zhang, S., Preacher, K. J., & Rucker, D. D. (2002). On the practice of dichotomization of quantitative variables. *Psychol Methods*, *7*, 19–40. doi: 10.1037//1082-989X.7.1.19
- Moratti, S., Clementz, B., Gao, Y., & Keil, A. (2007). Neural mechanisms of evoked oscillations: stability and interaction with transient events. *Hum Brain Mapp*, *28*, 1318–1333. doi: 10.1002/hbm.20342
- Murphy, K., Roodenrys, S., & Fox, A. (2006). Event-related potential correlates of the word length effect in working memory. *Brain Res*, *1112*, 179–190. doi: 10.1016/j.brainres.2006.07.007
- Newman, A., Tremblay, A., Nichols, E. S., Neville, H. J., & Ullman, M. (2012). The influence of language proficiency on lexical semantic processing in native and late learners of english. *J Cognitive Neurosci*, *25*, 1205–1223. doi: 10.1162/jocn_a_00143
- Oldfield, R. C. (1971). The assessment and analysis of handedness: The edinburgh inventory. *Neuropsychologica*, *9*, 97–113. doi: 10.1016/0028-3932(71)90067-4
- Pablos Martin, X., Deltenre, P., Rossion, B., Hoonhorst, I., & Colin, C. (2007). Perceptual biases for rhythm: The Mismatch Negativity latency indexes the privileged status of binary vs non-binary interval ratios. *Clin Neurophysiol*, *118*, 2709–2715. doi: 10.1016/j.clinph.2007.08.019
- Penolazzi, B., Hauk, O., & Pulvermüller, F. (2007). Early semantic context integration and lexical access as revealed by event-related brain potentials. *Biol Psychol*, *72*, 373–388. doi: 10.1016/j.biopsycho.2006.09.008
- Pinheiro, J. C., & Bates, D. M. (2000). *Mixed-effects models in s and s-plus*. New York: Springer.
- Pritchett, S., Zilberg, E., Xu, Z., Myles, P., Brown, I., & Burton, D. (2010). Peak and averaged bicoherence for different eeg patterns during general anaesthesia. *Biomed Eng Online*, *9*, 9–76. doi: 10.1186/1475-925X-9-76
- Quené, H., & van den Bergh, H. (2008). Examples of mixed-effects modeling with crossed random effects and with binomial data. *J Mem Lang*, *59*, 413–425. doi: 10.1016/j.jml.2008.02.002

- R Development Core Team. (2013). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria. Retrieved from <http://www.R-project.org/> (ISBN 3-900051-07-0)
- Rupert, D., Wand, M., & Carroll, R. (2003). *Semiparametric Regression*. Cambridge: Cambridge University Press.
- Sahai, H., & Ageel, M. I. (1997). *Analysis of Variance: Fixed, Random and Mixed Models*. Boston: Birkhauser.
- Sereno, S., Rayner, K., & Posner, M. (1998). Establishing a time-line of word recognition: evidence from eye movements and event-related potentials. *NeuroReport*, 9(10), 2195–2200. doi: 10.1097/00001756-199807130-00009
- Tremblay, A. (2009). *Processing advantages of lexical bundles: Evidence from self-paced reading, word ad sentence recall, and free recall with event-related brain potential recordings*. (PhD Dissertation). University of Alberta, Edmonton, Canada.
- Tremblay, A., & Baayen, R. H. (2010). Holistic processing of regular four-word sequences: A behavioral and ERP study of the effects of structure, frequency, and probability on immediate free recall. In D. Wood (Ed.), *Perspectives on Formulaic Language: Acquisition and Communication* (pp. 151–173). London and New York: Continuum.
- Vossen, H., van Breukelen, G., Hermens, H., van Os, J., & Lousberg, R. (2011). More potential in statistical analyses of event-related potentials: a mixed regression approach. *Int J Method Psych*, 20, e56–e68. doi: 10.1002/mpr.348
- Webb, S. J., Long, J. D., & Nelson, C. A. (2005). A longitudinal investigation of visual event-related potentials in the first year of life. *Developmental Sci*, 8, 605–616. doi: 10.1111/j.1467-7687.2005.00452.x
- Wierda, S., van Rijn, H., Taatgen, N., & Martens, S. (2010). Distracting the mind improves performance: An erp study. *PloS One*, 5, e15024. doi: 10.1371/journal.pone.0015024
- Wood, S. N. (2006). *Generalized Additive Models*. New York: Chapman & Hall/CRC.
- Wood, S. N. (2012). mgcv: Mixed GAM Computation Vehicle with GCV/AIC/REML Smoothness Estimation. [Computer software manual]. Retrieved from <http://CRAN.R-project.org/package=mgcv> (R package version 1.7-22)
- Wu, H., & Zhang, J. (2006). *Nonparametric regression methods for longitudinal data analysis*. Hoboken, NJ: Wiley.
- Wu, L. (2010). *Mixed Effects Models for Complex Data*. New York: Chapman & Hall.
- Zuur, A. F., Ieno, E. N., Walker, N. J., Saveliev, A. A., & Smith, G. M. (2009). *Mixed effects models and extensions in ecology with R*. New York: Springer.