INTERSPEECH 2016
September 8–12, 2016, San Francisco, USA

Comparing different methods for analyzing ERP signals

Kimberley Mulder[1], *Louis ten Bosch*[1], *Lou Boves*[1]

[1]Center for Language Studies, Radboud University, Nijmegen, the Netherlands
Kimberley.Mulder@let.ru.nl, l.tenbosch@let.ru.nl, l.boves@ru.nl

Abstract

Event-Related Potential (ERP) signals obtained from EEG recordings are widely used for studying cognitive processes in spoken language processing. The computation of ERPs involves averaging over multiple participants and multiple stimuli. Especially with speech stimuli, which also evoke substantial exogenous excitation, even averaging within conditions results in pooling many sources of variance. This raises questions about the statistical processing needed to uncover reliable differences between conditions. In this study we investigate differences between ERPs when participants listened to full and reduced pronunciations of verb forms in Dutch, in isolation and in mid-sentence position. Conventional statistical analysis uncovers some (but not all) differences between full and reduced forms in isolation, but not in mid-sentence position. In this paper, we show that linear mixed models (lmer) and generalized additive models (gam), which are able to account for participant- and stimulus-related variance may uncover more effects than conventional statistical models. However, depending on the complexity of the data, lmer and gam models may not be able to fit the data closely enough to warrant blind interpretation of the summary output. We discuss opportunities and threats of these approaches to analyzing ERP signals.

Index Terms: ERP based on EEG, reduced vs. unreduced speech, statistical analysis, generalized additive models

1. Introduction

Speech comprehension involves a complex sequence of processes for accessing information at different levels of representation. To study listening comprehension, one should be able to separate and account for the different cognitive processes that are involved during the unfolding of an auditory stimulus. Before the advent of sophisticated neurophysiological experimental techniques, studies on speech processing were mainly restricted to collecting button presses (i.e., that reflect response latencies or accuracy judgements) to isolated words or sentences. These experimental techniques provide output measures that are the sum of multiple cognitive (stimulus-related and stimulus-unrelated) processes that affect the processing of a stimulus. For instance, a response latency for a given stimulus not only includes the time it takes to actually process the stimulus, but also the decision about the response to be given, the motor planning and execution (pressing a specific button). This makes most behavioral experimental techniques unsuitable to study the processing of a stimulus proper, because they cannot separate the individual cognitive processes. Moreover, output measures do not capture the continuous nature of the speech signal and cannot account for the time course of effects on the processing of auditory stimuli (i.e., how different variables affect the processing of a stimulus as the signal unfolds).

One way for studying the time course of auditory processing is to use Event-Related Potentials (ERPSs). ERPs provide a millisecond-to-millisecond record of the electrical changes evoked by an ongoing stimulus. With this technique, one can capture the continuous nature of the speech signal. However, ERPs are averages over multiple responses, almost always of multiple participants and/or multiple stimuli. Therefore, the averaging process is (perhaps tacitly) based on the assumption that the underlying processes are very similar in all participants and all stimuli. These assumptions justify the use of repeated measure ANOVA techniques for comparing average amplitudes of the ERP signals between experimental conditions. It is not unusual for those ANOVAs to fail to return significant differences between conditions that researchers expected to differ substantially. This has raised the question whether the assumptions underlying ANOVA do indeed hold, and whether more advanced analysis techniques can uncover effects that are not uncovered by ANOVAs. This paper addresses this question on the basis of an experiment aimed to investigate whether native listeners of Dutch process full and reduced verb forms differently.

2. Description of the experiment

Right-handed undergraduate students passively listened to full and reduced pronunciations of verb forms in three conditions: the words presented in isolation (N = 31, mean age = 21.3, sd age = 2.7), the words in mid-sentence (N = 28, mean age = 20.8, sd age = 2.0) and in sentence-final position (N = 27, mean age = 21.7, sd age = 1.8) [1, 2]. In this paper we only discuss the first two conditions. Stimuli were presented over Sennheiser HD215 headphones using the Presentation software (Neurobehavioural Systems, www.neurobs.com). Participants were instructed to attentively listen and were told that they would get questions about the words and sentences they were about to hear. The goal of the experiment was to test whether full forms have an advantage over reduced forms in all conditions. That advantage should show up in differences between two ERP components, an N100 related to acoustic processing and an N400 related to semantic integration.

2.1. Stimulus materials

We selected 80 Dutch verb forms starting with the unstressed prefixes *be-* (/bə/, e.g., *bevallen* /bəˈvalə/, 'to give birth'), *ge-* (/xə/, e.g., *genieten* /xəˈniːtə/, 'to enjoy'), or *ver-* (/vər/, e.g., *vertellen* /vərˈtɛlə/, 'to tell'). When pronounced in their full forms, these prefixes all contain a schwa. We only selected those verb forms whose second syllable start with a consonant. Out of the 80 verb forms, 31 with *ver-*, 31 with *be-* and 18 with *ge-*. In addition, we selected 120 filler verb forms in order to make the stimuli better represent the Dutch lexicon, which contains many verb forms that do not start with one of the three prefixes.

In mid-sentence position, the verb forms were the main

http://dx.doi.org/10.21437/Interspeech.2016-967

verb in 30 sentences (e.g., *De bewoners bereiden de open dag voor*, 'The inhabitants prepare the open day'); were preceded by an auxiliary verb in 26 sentences (e.g., *Hij wil alles verdelen over zijn kleinkinderen* 'He wants everything to divide over his grandchildren',) or were part of the subject in 24 sentences (e.g., *Het jongetje genezen was voor deze arts niet moeilijk* 'The boy curing was for this doctor not hard'). The different syntactic functions ensure that our results do not depend on a specific syntactic construction. The target verb form was always preceded by four syllables. The words to be presented in isolation were segmented from sentence-final position. These forms were always preceded by an auxiliary. For all sentences, sentence accent was never on the target verb form or the preceding syllable. The semantic context up until the target verb form was kept as neutral as possible.

Sentences were recorded by a male native speaker of Dutch three times: Once without specific instructions, once with the instruction to pronounce all verb forms in full, and once with the instruction to pronounce the verb forms without the prefixal schwas. For the filler sentences, there was no specific instruction. In the sentences that the speaker produced without having received any instructions (i.e., the 'carrier' sentence), prefixal schwa was present in 68.4% of the verb forms and absent or unclear in the 31.6% of the verb forms that occurred in mid-sentence position, and present in 52.5% and absent or unclear in 47.5% of the verb forms of the carrier sentences that were used for the words to be presented in isolation.

The reduced and unreduced verb forms were spliced out of their original sentences and were pasted into the carrier sentence (in mid-sentence position) or presented in isolation. This was done to make sure that the reduced and unreduced sentences only differed with respect to the realization of the target verb form. The spliced reduced and full verb forms had a mean schwa duration of 3 ms and 42 ms in mid-sentence, and of 0 ms and 43 ms in isolation, respectively. The mean durations of the entire word were 430 ms for the reduced forms and 495 ms for the full forms in mid-sentence position, and 739 ms and 782 ms in isolation, respectively.

2.1.1. Data acquisition

The EEG signal was recorded with 26 active electrodes mounted in an elastic cap (Acticap). Electrode positions were a subset of the international $10-20$ system, consisting of four midline electrodes (Fz, Cz, Pz, and Oz) and 22 lateral electrodes (Fp1/Fp2, F3/F4, F7/F8, FC1/FC2, FC5/FC6, C3/C4, T7/T8, CP1/CP2, CP5/CP6, P3/P4, and P7/P8). Moreover, an electrode was placed on each of the mastoids and each electrode was referenced online to the left mastoid. The electro-oculogram (EOG) was recorded by two vertical electrodes placed above and below the right eye and by two horizontal electrodes with a right to left canthal montage.

Electrode impedance was kept below $5k\Omega$. The EEG and EOG signals were amplified (band pass = $0.02-100$ Hz), and digitized with a sampling frequency of 500 Hz. Before data analysis, the signal was re-referenced to the average of the left and right mastoids and digitally filtered with a high cut-off filter of 30 Hz. Next, the continuous EEG was segmented into stimulus-time-locked epochs, starting from 200 ms before target word onset up to 800 ms after onset. Ocular artifacts were identified and removed with independent component analysis (ICA). After ICA, single trials that still contained artifacts were removed by a semi-automatic rejection routine. The period of 200 ms preceding the target word onset was used for baseline correction.

In this paper we confine ourselves to analyzing a single EEG signal, viz. Cz.

3. Isolated Words

Behavioral studies have shown that, at least in isolation, full forms benefit from a processing advantage over reduced forms (e.g. [3]). The ERP data from the isolated words for electrode Cz are shown in Figure 1. The green curve in the upper panel of Figure 1 shows the result of a sample-by-sample $t-$test on the average ERP signals from the full and reduced words. In a way, this corresponds to an ANOVA on very narrow bins, each containing a single sample (c.f. [4]). From the lower panel of Figure 1 it can be seen that there is a small processing advantage for full forms presented in isolation. More negative N100 amplitudes were observed for reduced forms (full red lines) relative to full forms (full blue lines). As speakers do not typically reduce words when uttered in isolation, the reduced forms could have been unexpected and, as a consequence, attracted more attention. A more negative N100 might also indicate speech segmentation problems[5], because absence of schwa in some reduced forms resulted in consonant clusters that only occur cross-word. However, as can be seen from the green curve in the upper panel of Figure 1, the amplitude difference between the ERPs for the full and reduced forms did not approached significance in the N100 region.

The green curve in the upper panel of Figure 1 suggests that the amplitudes of the ERP for the full and reduced forms differ significantly in the time interval from 300 to 400 ms. We interpret this as a difference in N400 peak latency for reduced and full forms, with a later N400 peak for reduced forms compared to full forms. As the N400 indexes lexical-semantic activation [6], this suggests that the activation of semantic representations is delayed for reduced forms. This is in line with behavioral findings that suggest that it takes more time to activate the semantic network for reduced forms [7].

For reduced forms, the N400 was preceded by a small negative peak around 270 ms, which was present at all central, fronto-central and centro-partietal sites and absent or less pronounced for full forms. We consider this peak to be the N250 [8], which might reflect a lexical selection process that occurs at the interface of lexical form and contextual meaning. Based on the incoming acoustic input, potentially matching lexical candidates are activated (e.g., [9]). [8] argue that an N250 effect would arise if contextual specifications do not support the activation of a lexical candidate, relative to a situation in which form-based activation is supported by the specifications of the context. The fact that words tend not to be reduced when they are pronounced in isolation therefore explains the finding that only reduced and not full forms generated a (clear) N250.

In sum, the ERP data seem to confirm a processing advantage for full forms presented in isolation. While a conventional ANOVA confirmed the expected difference related to the N400 component, it failed to uncover the – also expected – N100 component.

3.1. Linear Mixed Effects Models

The failure to find a significant N100 effect may be due to several different causes. Maybe the most obvious cause is the true absence of an effect, which would suggest that full and reduced versions of the verbs heard as words spoken in isolation are equally likely and equally natural in normal speech. However,

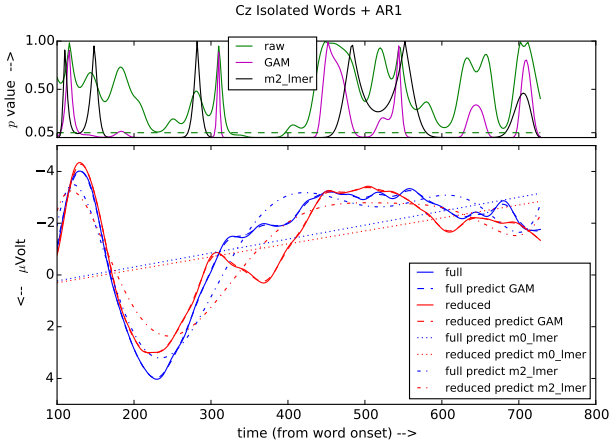

Figure 1: *Event-related potentials (ERPs) to full forms and reduced forms presented in isolation for electrode Cz. Negativity is plotted upwards. Top panel: by-sample t-tests. X-axis represents time in ms. For details, see text.*

we cannot exclude the possibility that there is some effect, but that it is obscured by pooling multiple sources of variance.

One way for investigating the contributions of several sources of variance is to replace ANOVA by a linear mixed effects model. For that purpose we used the R-package `lme4` [10]. We started with a fairly simple model:

$$m0\_lmer \leftarrow lmer(amplitude \sim type\_fac + type2\_fac + word\_dur + time + (1|subject\_fac) + (1|stimulus\_fac), data = dat)[-1mm]$$

The output of that model is summarized in Table 1. The predictor *type_fac* corresponds to the planned difference between full and reduced forms; the predictor *type2_fac* has the three levels *be-, ge-, ver-*. We included *time* as a predictor, to prevent the model from collapsing all samples. We included subject and stimulus as random factors, because we aimed to remove as much participant- and stimulus-related variance as possible. From Table 1 it can be seen that the factor full/reduced is significant, as are the duration of the stimulus words and the factor *time*. However, from the summary output it is not possible to know in what time interval the ERP amplitudes of the full and reduced forms differ significantly. We hoped to be able to glance that information from the predictions of the model of the amplitude of the ERPs of the individual stimuli. The averages of these predictions are shown by the dotted lines in the lower panel of Figure 1. However, these approximations turn out to be straight lines. Using hindsight, this does not come as a surprise: ERP signals are not linear functions of time, and as a result a strictly linear model cannot achieve close approximations.

An analysis without the factor *word_dur* returns only *ver-* and *time* as significant. We believe that this is due to the fact

Table 1: Summary of the `m0_lmer` model.

	Estimate	Std. Error	*t* value
(Intercept)	-1.6022229	0.6161846	-2.60
type_fac2	0.5055082	0.0508597	9.94
type2_fac2	-0.0402940	0.4675198	-0.09
type2_fac3	0.9116503	0.3964945	2.30
word_dur	0.0079978	0.0005186	15.42
time	-0.0116187	0.0001587	-73.23

Table 2: Partial summary of the `m2_lmer` model.

	Estimate	Std. Error	t value
(Intercept)	-5.335e+00	6.427e-01	-8.30
type2_fac2	4.663e-02	4.762e-01	0.10
type2_fac3	9.478e-01	4.039e-01	2.35
word_dur	8.363e-03	5.110e-04	16.37
type_fac2	5.835e-01	3.797e-01	1.54
⋮	⋮	⋮	⋮

that the model must have at least basic information information about the temporal structure of the acoustic stimuli to be able to account for possible time shifts in the neural processes related to different stimulus durations.

3.1.1. Adding time as a non-linear predictor

The `lmer()` function in *R* offers the possibility of adding higher-order polynomial functions of predictor variables to the model. Because we have seen that ERP amplitude is not a linear function of time, we experimented with adding polynomials of time with increasing order. Table 2 shows a small part of the output of model `m2_lmer`.

$$m2\_lmer \leftarrow lmer(amplitude \sim type2\_fac + word\_dur + type\_fac * poly(time2, 7) + (1 + type\_fac|subject\_fac) + (1|stimulus\_fac), data = dat)$$

It can be seen that the factor `type_fac` is no longer significant in its own right. However, many of its interactions with polynomials of *time* are highly significant (not shown).

The dash-dot lines in the lower panel of Figure 1 show the approximation of the average ERP signals for full and reduced versions of the verb forms obtained with the model `m2_lmer`. The black curve in the upper panel shows the p values from sample-by-sample t-tests for the difference between the approximated ERP signals. It can be seen that the model suggests that the amplitudes of the ERP signals pertaining to the full and reduced words differ significantly in several long, connected time intervals. However, from the lower panel of Figure 1 is is evident that the approximated ERP signals do not fit the raw signals very well. Maybe it should not come as a surprise that the approximations of the model do not reproduce the high-frequency oscillations in the raw averages; also, it is not clear whether these oscillations correspond to relevant cognitive processes. However, it is alarming that the approximation of the reduced ERP does not reproduce the N250 component, to which we did assign a cognitive interpretation. This raises questions about the validity of model `m2_lmer`.

3.2. Generalized Additive Models

Adding polynomials of the predictor *time* to an `lmer()` model is a kludge, if only because predictions outside the time interval in the analysis will very likely be completely wrong. The recently introduced family of Generalized Additive Models (GAM) offers an attractive alternative, by replacing polynomials by regression splines (see [11] for an introduction to GAM analysis). GAM determines a linear and/or non-linear equation that strikes a balance between over-fitting and overgeneralizing a set of data through a process called penalized iteratively reweighted least squares. The algorithm separates the parametric

from the nonparametric part of the fit, and fits the parametric part using weighted linear least squares.

[12, 13] applied GAMs to analyze EEG signals, using the *R*-package `mgcv` [14]. We used the same software to obtain the GAM model `gam2`.

$$\begin{aligned} gam2 \leftarrow bam(amplitude \sim type\_fac + type2\_fac \\ +s(word_dur) + s(time, by = type_fac, k = 50) \\ +ti(time, schwa_dur) + s(subject_fac, bs = "re") \\ +s(stimulus_fac, bs = "re"), data = dat, \\ samfrac = 0.1, gc.level = 2, correlation = corAR1()) \end{aligned}$$

The approximations of the raw ERP signals obtained with this model are shown as dashed lines in the lower panel of Figure 1. If the dashed lines are extremely difficult to distinguish from the full lines, that is because the approximation with `gam2` is close to perfect. That is not to say that the approximations of individual EEG signals is very good. `gam2` only accounts for slightly less than 10% of the total variance in the data.

The p-values of a sample-by-sample t-test on the approximated signals again shows substantial connected time intervals in which the ERP signals of the full and reduced words, with at least part of the participant- and item-related variance removed, differ significantly. Specifically, in the GAM analysis the difference in the N100 region appears to be significant. Because of the close approximation of the average ERP signals by the GAM model we trust that here the p-values are meaningful.

4. Mid-sentence data

In [2] it was found that in mid-sentence position conventional ANOVAs did not discover differences between the ERP signals of full and reduced forms. This finding was explained by the fact that in mid-sentence position reduced pronunciations are at least as natural as full forms. However, it is worthwhile investigating whether significant differences can be found if part of the subject- and item-related variance is removed. The results of an `lmer()` and a `gam` model are shown in Figure 2. The full green line in the top panel confirms that t-tests do not detect significant differences in the raw ERPs.

The sample-by-sample t-tests on the approximations by the `lmer` and `gam` models suggest that the full and reduced ERP signals differ significantly over several connected time intervals. However, from the lower panel in Figure 2 it is evident that both model approximations differ considerably from the raw averages of the full and reduced ERP signals. It can be argued that these averages are so complex (have so many local minima and maxima) that both models fail to capture the structure if any-in the data. The adjusted R^2 of the `gam2` model applied to the mid-sentence is only 4.6%. It seems that the EEG signals corresponding to the words in mid-sentence position contain a substantial amount of variance that is not related to the predictors used in the model. A probable source is the exogenous excitation due to the speech preceding the interval under analysis.

Heeding the lessons learned from the analyzes of the isolated word data, we refrain from interpreting the differences that are significant according to the `lmer` and `gam` models. Only the interval around 150 ms after the onset of the crucial word, where the analysis of the raw signals approaches significance, is likely to be meaningful. The larger negative-going amplitude of the full forms suggests that in mid-sentence position reduced forms are more likely.

In [15] it is recommended to evaluate the quality of `gam` models by analyzing the autocorrelation in the prediction error. Although this may work well when predicting time series of

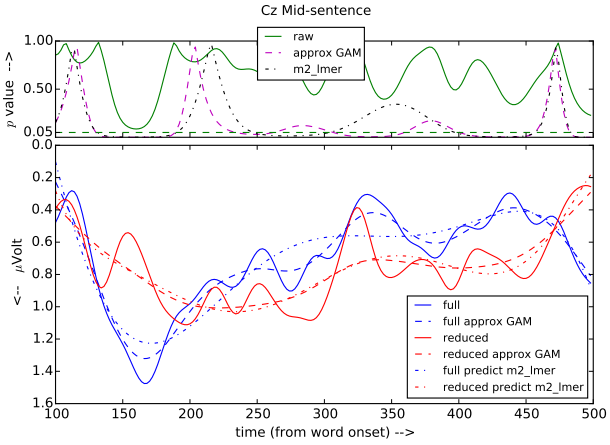

Figure 2: *Event-related potentials (ERPs) to full forms and reduced forms presented in mid-sentence position for electrode Cz. Negativity is plotted upwards. Top panel: by-sample t-tests. For details, see text.*

reaction times, which are indeed characterized by a first-order auto-regressive process that should -and can- be accounted for by a linear mixed effects model [16, 17], the situation with EEG signals is very different. Figures 1 and 2 show that the average ERP signals related to full and reduced pronunciations are much more complex than a first-order auto-regressive process. Especially for the mid-sentence data there is a need for additional processing to remove the exogenous excitations. The autocorrelation that is naturally present in EEG signals can only be removed by a model that accurately predicts all individual signals. No linear model with the predictors that we have available can be expected to accomplish this. We plan to investigate whether the linear deconvolution approach proposed in [18, 19] can remove enough exogenous excitation from the EEG signals to yield average ERP signals that come closer to a sequence of established ERP components.

5. Conclusions

In this paper we investigated several statistical procedures for analyzing EEG/ERP signals. We want to use such models to reduce the subject- and item-dependent variance sources that may obscure true differences between conditions, such as listening to full and reduced pronunciations of words. We found that linear models can only begin to approximate EEG signals if time can be included as a non-linear predictor. In `lmer` models this can be accomplished by introducing `poly(time, n)`, where n specifies the order of the polynomial. Although such `lmer` models suggested that there are substantial connected time intervals where full and reduced pronunciations differ, we concluded that the fit of the model to the data was not good enough to be able to interpret the results. For the isolated words the `gam` model did obtain a very close approximation. However, the mid-sentence signals appear to contain so much variation that is not related to the predictors that even a gam model fit is not good enough.

6. Acknowledgements

This work was supported by an ERC consolidator grant awarded to prof. Mirjam Ernestus (nr 284108).

7. References

[1] L. Drijvers, K. Mulder, and M. Ernestus, "Alpha and gamma band oscillations index differential processing of acoustically reduced and full forms," *Brain and Language*, vol. 153-154, pp. 27–37, 2016.

[2] K. Mulder, L. Drijvers, and M. Ernestus, "The time course in processing reduced and unreduced word pronunciation variants: An ERP study," submitted.

[3] M. Ernestus and H. Baayen, "Paradigmatic effects in auditory word recognition: The case of alternating voice in Dutch," *Language and Cognitive Processes*, vol. 22, no. 1, pp. 1–24, 2007.

[4] E. Maris and R. Oostenveld, "Nonparametric statistical testing of eeg- and meg-data," *Journal of Neuroscience Methods*, vol. 164, no. 1, pp. 177 – 190, 2007.

[5] L. D. Sanders and H. J. Neville, "An {ERP} study of continuous speech processing: Ii. segmentation, semantics, and syntax in non-native speakers," *Cognitive Brain Research*, vol. 15, no. 3, pp. 214 – 227, 2003.

[6] M. Kutas and K. D. Federmeier, "Thirty years and counting: Finding meaning in the N400 component of the event-related brain potential (ERP)," *Annual Review of Psychology*, vol. 62, pp. 621–647, 2011.

[7] M. Van de Ven, B. V. Tucker, and M. Ernestus, "Semantic context effects in the comprehension of reduced pronunciation variants," *Memory & Cognition*, vol. 39, no. 7, pp. 1301–1316, 2011.

[8] P. Hagoort and C. Brown, "ERP effects of listening to speech: Semantic erp effects." *Neuropsychologia*, vol. 38, pp. 1518–1530, 2000.

[9] W. Marslen-Wilson and A. Welsh, "Processing interactions and lexical access during word recognition in continuous speech," *Cognitive Psychology*, vol. 10, pp. 29–63, 1978.

[10] D. Bates, M. Mächler, B. Bolker, and S. Walker, "Fitting linear mixed-effects models using lme4," *Journal of Statistical Software*, vol. 67, no. 1, pp. 1–48, 2015.

[11] S. N. Wood, *Generalized additive models*. New York: Chapman & Hall/CRC, 2006.

[12] A. Tremblay and R. H. Baayen, "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 *Perspectives on formulaic language: Acquisition and communication*, D. Wood, Ed. London: The Continuum International Publishing Group, 2010, pp. 151–173.

[13] T. Kryuchkova, B. Tucker, L. Wurm, and R. H. Baayen, "Danger and usefulness are detected early in auditory lexical processing: evidence from electroencephalography," *Brain and Language*, vol. 122, pp. 81–91, 2012.

[14] S. Wood, *Generalized Additive Models: An Introduction with R*. Chapman & Hall/CRC, 2006.

[15] R. H. Baayen, J. van Rij, C. de Cat, and S. N. Wood, "Autocorrelated errors in experimental data in the language sciences: Some solutions offered by Generalized Additive Mixed Models," in *Mixed Effects Regression Models in Linguistics*, D. Speelman, K. Heylen, and D. Geeraerts, Eds. Berlin: Springer, to appear).

[16] L. ten Bosch, L. Boves, and M. Ernestus, "Comparing reaction time sequences from human participants and computational models," in *Proceedings of Interspeech*, Singapore, 2014.

[17] ——, "DIANA: towards computational modeling reaction times in lexical decision in north american english," in *Proceedings of Interspeech*, Dresden, 2015.

[18] E. Lalor and J. Foxe, "Neural responses to uninterrupted natural speech can be extracted with precise temporal resolution," *European Journal of Neuroscience*, vol. 31, no. 1, pp. 189–193, 2010.

[19] N. Gonçalves, R. Whelan, J. Foxe, and E. Lalor, "Towards obtaining spatiotemporally precise responses to continuous sensory stimuli in humans: a general linear modeling approach to EEG," *NeuroImage*, vol. 97, pp. 196–205, 2014.