Combining data-oriented and process-oriented approaches to modeling reaction time data

L. ten Bosch\textsuperscript{1}, L. Boves\textsuperscript{1}, M. Ernestus\textsuperscript{1,2}

\textsuperscript{1}Radboud University Nijmegen
\textsuperscript{2}Max Planck Institute for Psycholinguistics
{l.tenbosch, l.boves}@let.ru.nl, m.ernestus@let.ru.nl

Abstract

This paper combines two different approaches to modeling reaction time data from lexical decision experiments, viz. a data-oriented statistical analysis by means of a linear mixed effects model, and a process-oriented computational model of human speech comprehension.

The linear mixed effect model is implemented by \texttt{lmer} in \texttt{R}. As computational model we apply DIANA, an end-to-end computational model which aims at modeling the cognitive processes underlying speech comprehension. DIANA takes as input the speech signal, and provides as output the orthographic transcriptions of the stimulus, a word/non-word judgment and the associated reaction time. Previous studies have shown that DIANA shows good results for large-scale lexical decision experiments in Dutch and North-American English.

We investigate whether predictors that appear significant in an \texttt{lmer} analysis and processes implemented in DIANA can be related and inform both approaches. Predictors such as ‘previous reaction time’ can be related to a process description; other predictors, such as ‘lexical neighborhood’ are hard-coded in \texttt{lmer} and emergent in DIANA. The analysis focuses on the interaction between subject variables and task variables in \texttt{lmer}, and the ways in which these interactions can be implemented in DIANA.

Index Terms: reaction times, local speed, computational modeling, spoken word recognition, data-modeling versus process modeling.

1. Introduction

The advent of ever larger compute power and data sets has intensified the simmering debate between proponents of data modeling or process modeling in all sciences, including psycholinguistics \cite{1}. Behavioral measures such as lexicality judgments and reaction times \cite{2, 3} can be analyzed with linear mixed models, exemplified by \texttt{lmer} in \texttt{R}, but also with process models, such as DIANA \cite{5, 6, 7} that can be seen as an elaboration of older models such as Shortlist-B \cite{8} or SpeM \cite{9}. Regression models ‘explain’ the structure in data sets on the basis of a number of predictors that are selected by the modeler. Process models ‘explain’—usually small dedicated—data sets by simulating a number of processes that the modeler assumes are going on in the brain. Obviously, data and process modeling reflect deeply diverging approaches. Still, we believe that the two types of model can and must inform each other. Data models can only truly advance our understanding if the predictors can in some way linked to plausible processes. Process models are much easier to develop if we have suggestions about plausible processes that could be derived from data models.

2. DIANA

DIANA aims to simulate participants’ behavior in experiments in spoken word comprehension. The model consists of three components: an Activation Component, a Decision Component, and an Execution Component (cf. Figure 1). Activation and Decision operate in parallel; once the Decision Component has made a decision, the Execution component is initiated. Importantly, like its predecessors SpeM and Shortlist-B DIANA was designed for simulating the processing of stimuli in isolation. Being a computational model of the cognitive processes involved in spoken word comprehension, DIANA does not simulate humanly effects such as waning attention or fatigue.

Activation Component

The Activation Component uses the signal to compute time-varying activations for all words in the lexicon. The current implementation handle lexicons of 40,000 entries. Entries in the lexicon are phonetically specified as (possibly several parallel) sequences of phone symbols. Each entry is accompanied by a prior probability, derived from the relative frequency of the word in a text corpus. The word activations are determined each 10ms by combining acoustic bottom-up information and top-down information given the word priors:

$$\log P(\text{signal|word}) + \lambda \log (P(\text{word}))$$

in which the model parameter $\lambda$ governs the balance between the bottom-up acoustic information (first term) and the top-down linguistic information (second term).

DIANA’s Activation Component performs two types of decoding. A lexical decoding computes the activations over time

\footnote{http://aphl.artsrn.ualberta.ca/?p=517}
pairs of words with varying degrees of reduction. The decision on the second occurrence of a word (the target) may be facilitated by the first (the prime). An ‘exemplar effect’ takes place when the RT on a target word is shorter when it is preceded by an acoustically similar prime than when it is preceded by an acoustically less similar prime.

The stimuli consisted of an equal number of Dutch words and pseudo-words; all were tri-syllabic infinitives, starting with an unstressed syllable (prefix) with a schwa (e.g., *beschrijven* ‘to describe’, *vertolken* ‘to interpret’). The pseudo-infinitives did not contain phonotactically illegal phoneme sequences. Stimuli were produced by one male (spk A) and one female speaker (spk B). All primes were existing infinitives. Primes and targets could differ in their degree of reduction; they were chosen to have different realizations of the same word. 48 native participants listened to 288 trials in which 34% of the trials formed word repetitions (similar to [15]). Each participant only heard one speaker.

### 4. Simulation 1

Lexical decision involves a potentially large number of partly independent and partly interacting processes. A mixed effects model that predicts the log(RT) sequences in such an experiment might look like:

$$\log \text{RT} = P_{\text{stim},1} + \ldots + P_{\text{stim},N} + P_{\text{subj},1} + \ldots + P_{\text{subj},M} + \text{linear interactions} + \text{non-linear terms} + \text{non-linear interactions} + \text{noise} + P_{\text{previous decisions}}$$

in which $P_{\text{stim}}$ denotes a predictor, $N$ and $M$ denote the number of stimulus-specific and subject-specific predictors; the last three terms are a shorthand for all interactions, non-linear terms, perceptual and decision noise ([16]), and parameters related to previous decisions (e.g. ‘previous RT’). It is reasonable to assume that processes that play a role early in the decision will be involved in many interactions; processes that come later may have fewer interactions, but it is questionable whether a regression model is at all capable of modeling the effect of the temporal order in which predictors operate.

Robust estimation of all (non)-linear interactions in Eq. 3 would require vastly more data than is available in psycholinguistic experiments. In practice, all non-linear terms are discarded, so that we end up with a linear mixed effects model with subjects and word as random variables. This resulted in an *lmer* model with speaker (male, female), word frequency, the prefix, the prime-target match/mismatch, the duration of the stimulus, and the reaction time given to the prime as fixed effects, and subject and word as random effects [10]. All non-significant effects and random slopes were excluded:

```r
modell = lmer(logRT ~ speaker + logwordfreqcgn + prefix + variant_match + logdurms + logRTprime + logRTprev + (1|subject) + (1 + logRTprev|word))
```

The output of this model (only the fixed factors) is shown in Table 1.

The question now arises which of the predictors in such an *lmer* model might correspond to processes in a model such as DIANA. We distinguish three types of predictors. The first type, which is most clearly related to a cognitive process, is the use of the “previous RT” (logRTprev) in the prediction of the current RT. However, DIANA processes all stimuli independently. Still, below it will appear that there is an interesting link.

---

**Figure 1:** Overview of DIANA. The model consists of three interrelated components: (1) an Activation Component that takes speech as input; its output is a weighted lattice of hypotheses, evolving over time (2) a Decision Component, which outputs the recognized word/non-word item and an estimated reaction time (3) an Execution Component which models the time it takes from the mental decision until the eventual overt reaction time (e.g. pressing a button).
The output of \( \text{lmer} \) with \( \text{logRTprevious} \) and \( \text{logRTprev} \) is used as a predictor; in \( \text{Diana} \) word frequency serves as a prior probability during the search \((17)\). The actual effect of word frequency on \( \text{Diana} \)’s RT depends also on the neighborhood density, which in turn depends on the composition of the lexicon. In \( \text{lmer} \) model neighborhood density must be specified as just another predictor.

The \( \text{lmer} \) model \( \text{modell} \) shows that mismatch between prime and target leads to significantly longer RTs on the target than in the match condition. In order to model this “exemplar” effect, \( \text{Diana} \) should be able to increase the prior probability of the full or reduced version of words as targets with the appropriate difference. Exactly how that should be done is not obvious. The \( \text{lmer} \) model \( \text{modell2} \), in which the difference between RTs on the prime and target is the dependent variable, shows why adapting the prior is complex:

\[
\text{model2} = \text{lmer}(\text{difflogRT} \sim \text{speaker} + \text{variant_match} + \logdurms + \logRTprev + (1|\text{subject}) + (1+\logRTprev|\text{word})
\]

The output of \( \text{model2} \), summarized in Table 2, shows that a mismatch between prime and target (reduced versus unreduced, or vice versa) leads to a significant difference of 0.024 in RT. Therefore, the wrapper around \( \text{Diana} \) should be able to adapt priors in the lexicon on the basis of the pronunciation \( \text{variant} \) that was recognized, not only on the basis of having the word type encountered previously. We simulated the RTs with \( \text{Diana} \), once with fixed priors in the lexicon, and once with priors that we increased as function of the prime-target mismatch. With fixed priors the correlation between the RTs predicted by \( \text{Diana} \) and the average RTs of the participants is 0.34, with adapted priors the correlation is raised to 0.38. This is a promising result, given the fact that the average correlation between RT sequences of pairs of participants is only 0.085. In this case, the improvement in \( \text{Diana} \) as a process model is motivated by the output of data-oriented regression modeling.

### 5. Simulation 2

In the second simulation we study the reverse, process modeling can inform regression approaches. It has been observed that the residual error in sequences of RTs in mixed effects models may contain a substantial autocorrelation \((18)\). That may even be the case if the model has the RT to the previous stimulus (RT previous), which is supposed to capture what is known as ‘local speed effects’ \((19)\) in RT sequences, as one of the predictors. In developing \( \text{Diana} \) maximizing the correlation between simulated RTs and average RTs of participants was an important criterion. Since \( \text{Diana} \) processes stimuli one by one and ignores the previous result, the comparison between RT sequences simulated by \( \text{Diana} \) and the observed human RT sequences only makes sense if the impact of the different ordering in which participants heard the stimuli is accounted for. This is done by detrending the human RT sequences by applying an autoregressive (AR) filter on the observed RT sequences. This filter generalizes the ‘previous RT’ to a weighted RT (denoted \( maRT = maRT[1, 2, \ldots, N] \) where \( N \) denotes the number of reaction time measurements in one session), which is based on all previous RTs with a weighting factor that is exponentially decaying with recency:

\[
maRT[i] = \alpha RT[i] + (1 - \alpha) maRT[i - 1]
\]

for \( i > 1 \), with \( 0 < \alpha \leq 1 \). The conventional ‘previous RT’ is obtained by setting \( \alpha = 1 \). After computing the maRT sequence, the difference \( RT[i] - maRT[i] \) is taken as the current detrended RT. This detrending removes both the long-term effects on the observed human RT sequence (health condition, age, gender, handedness, general cognitive abilities, etc. \((16)\) as well as the mid-term effects (attention fluctuation, strategy changes, learning effects, fatigue), leaving the stimulus-related aspects of interest (lexical status, morphological complexity, the density of its lexical neighborhood, frequency effects, indexical effects) \((e.g. \, [20, 21, 22, 23, 24, 25, 26] \) and references therein). From the literature, it is known that RTs of the previous 5 to 10 stimuli may influence the current RT \((26)\), which corresponds to a value of \( \alpha \) between 0.1 and 0.2.

Pairs of participants often show extremely low correlations in their RT sequences \((< 0.15, [6, 7])\). The same is true in the data under analysis here. This is caused by the influence of a multitude of participant dependent long-term and mid-term effects on the RT given to the current stimulus \((60)\). It appears that the between-participant correlation can be substantially increased by detrending RT sequences with an AR filter with \( \alpha \approx 0.1 \).

In this simulation, we first investigated the impact of the detrending parameter \( \alpha \) on the average correlation between the observed RT sequences across all pairs of subjects, as obtained in \((10)\). The results are given in Figure 2. The horizontal axis displays the range of \( \alpha \); the vertical axis shows the average correlation. The rightmost point \( \alpha = 1 \) coincides with the choice in the \( \text{lmer} \) model in \((10)\), presented in the previous section.

The figure shows that, in general, the average correlation increases with decreasing value of \( \alpha \). This holds for RT data in which participants only heard the male or female speaker.
(dashed lines), as well for the combined model in which data from both male and female speaker are combined (solid line). The optimum is reached for $\alpha \approx 0.05$, showing that in this lexical decision experiment the reaction times of about 20 previous trials play a role in the current RT. The same trend has now been observed for RT data in many different lexical decision experiments ([5, 7]). It clearly shows that the way in which the modeling of the local speed is dealt with may have a serious impact on the match between the RT sequences simulated by a model and obtained from participants.

Second, we studied the impact of $\alpha$ on the significance of the predictor match/mismatch. Figure 3 shows the value of the $t$ statistic as provided by lmer for the predictor match/mismatch, on which the existence of the exemplar effect in this priming experiment ([10]) hinges. The $t$ statistic is displayed along the vertical axis. The dashed lines show the $t$ statistic as function of the detrending parameter $\alpha$; the thin horizontal lines indicate the constant values independent of $\alpha$ as in the lmer model from [10]. This figure clearly shows the dependency of the significance of the match/mismatch condition in prediction of RT as a function of $\alpha$. One observes that the significance is also gender dependent. On the combined set, there remains a clear significance. The significance is larger than reported in [10] if $\alpha$ is chosen smaller than 0.2, because this value of $\alpha$ properly accounts for the fact that RTs for the male speaker are affected by a large number of preceding stimuli.

In this case, the lmer analysis is made more precise by using the DIANA-inspired detrending of the observed RT sequences to remove the local speed effect, before comparison takes place between the human RT sequences and the RT sequences simulated by DIANA.

6. Discussion and conclusion

This paper presents two simulations that show how data-oriented and process-oriented models can inform each other. In the first simulation, we show how a data-oriented lmer uncov-

Figure 2: The average correlation between the RT sequences across all pairs of subjects, as a function of the detrending parameter $\alpha$. 

Figure 3: Values of the $t$-statistic for the male (spk A, in blue, lower curves), and female speaker (B, in red, middle curves), as well as for both (upper curves in black). Values below 1.95 relate to insignificance of the match/mismatch condition in the prediction of RT.
8. References


