

SNOD – Autoregression models

Jens Roeser

June 24, 2019

Overview

The current version of this document presents an analysis of a random subset of 20 ppts / 20 images.

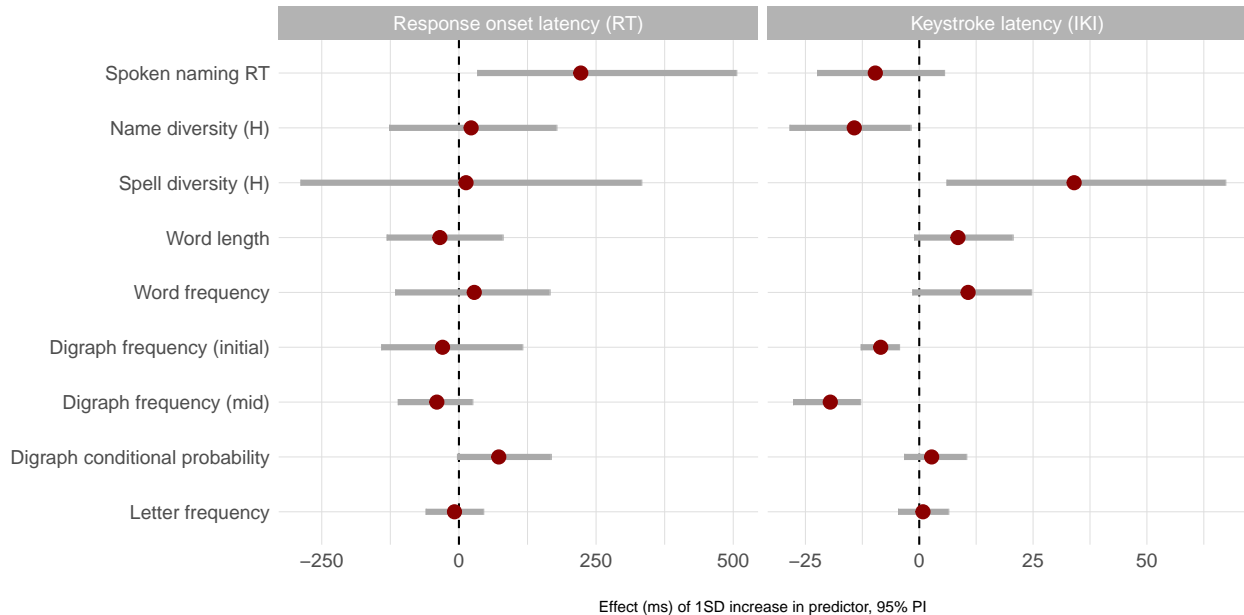
This report summarises the analysis of keystroke data from an image naming task using autoregression model to account for autocorrelations between subsequent keystrokes. All analyses were implemented in Stan. Random intercepts were included for participants and images. The autoregression between keystrokes is captured in the parameter ϕ . All models include an error variance σ . Models were fitted with weakly informative regulating priors.

Two general analyses are presented. The first analysis incorporates various predictors of image naming and writing (see below). A hierarchical model was fitted with hyperpriors on intercepts α and the different effects β separated for the first keystroke (the naming response time) and after keystroke onset (within-word IKIs).

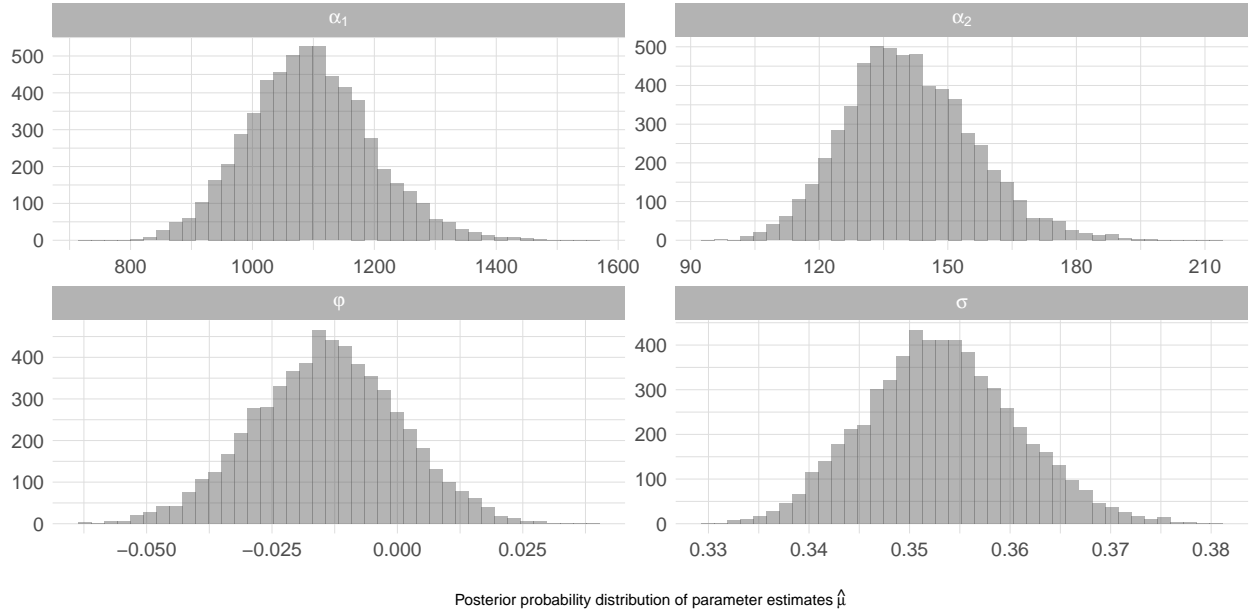
The second analysis tested an extension to mixture models. This extension to mixture models allows us to model effects on the naming process as probabilistic rather than deterministic. For simplicity this analysis focuses on the spelling H effect only.

Autoregression model predictors for written naming

First, we fitted a first order autoregression model with various predictors for written image naming (see AR1.stan). Intercepts α and effects β were estimated separately for before and after production onset keystrokes using hyperpriors on parameters before and after onset in a hierarchical model model with hyperpriors on both. Keystroke data were not trimmed. The autoregression between subsequent keystrokes was captured in ϕ . The following figure shows the estimated effects of spelling H separately for before onset response times and within-word keystroke intervals.



The distribution of the remaining parameter estimates is shown in the following figure. The intercept α are shown for the response time latency (1) and the within-word keystroke intervals (2). The autoregression is captured in the parameter ϕ and the error variance in the parameter σ .



Extension of the autoregression model to mixture models

The following series of autoregression models is focusing on the effect of spelling H effect. Model complexity was increased incrementally to test for the benefits of various extensions. The models are generally similar to the previous model. All models are first order autoregression models (autoregression component ϕ) with random intercepts for participants and images, lognormal distribution, and variance component σ . For simplicity the following models focus on the spelling H effect but can be extended to other variables.

Conceptual features of models presented from the simplest model to the most complex model:

- **Model 1:** no spelling H effect; intercept α and autoregression component ϕ [**ARnull.stan**]
- **Model 2:** no spelling H effect and different intercept α 1 and 2 for before and after onset IKIs, respectively (this was carried forward to the remaining models) [**ARnull2.stan**]
- **Model 3:** as Model 2 but with spelling H effect (across before/after onset) [**ARspell.stan**]
- **Model 4:** as Model 2 but with spelling H effect for before/after onset [**ARspell2.stan**]
- **Model 5:** as Model 4 but with different spelling effect for 0 spelling H indicated as d_spell [1] before onset, [2] after onset [**AR0spell.stan**]
- **Model 6:** as Model 4 but with mixture component for spelling H effect before onset (θ is mixing proportion indicating the probability of longer values; δ is the increased RT for those) [**ARspell2MoG.stan**]
- **Model 7:** as Model 4 but with mixture component after onset IKIs (θ is the mixing proportion for long values; δ is the increase in IKIs) [**ARspell2MoG3.stan**]
- **Model 8:** as Model 4 but with mixture components before [1] and after [2] onset (theta is the mixing proportion for long values; δ is the increase in RT/IKIs) [**ARspell2MoG2.stan**]

Model comparisons can be found in Table 1. The best fitting model is Model 7, the autoregression model with a mixture component for the IKIs after typing onset. These model comparisons show that that (a) keystroke intervals require a different intercept value for before and after production onset, (b) the spelling H effect has a different effect on these latencies before and after onset, and (c) the spelling H effect is to some extent probabilistic.

The estimated model parameter of the best fitting model are presented in Table 2. The best fitting model is

Table 1: Model comparisons. Models are ordered from the model with the highest predictive performance \widehat{elpd} to the model with the lowest predictive performance. $\Delta\widehat{elpd}$ indicates the difference between the best fitting model (first row) and the remaining models.

	$\Delta\widehat{elpd}$	ΔSE	\widehat{elpd}	SE
Model 7	0.00	0.00	-588.04	34.05
Model 8	-2.84	1.43	-590.88	34.41
Model 4	-76.94	18.17	-664.97	42.23
Model 2	-77.23	18.37	-665.27	42.32
Model 3	-77.36	18.23	-665.39	42.26
Model 5	-77.53	18.21	-665.56	42.23
Model 6	-151.41	34.64	-739.45	62.49
Model 1	-434.90	23.31	-1022.94	30.57

Table 2: Model parameters of the autoregression mixture model. Parameter value is indicated as its most probable parameter value $\hat{\mu}$ and 95% probability intervals. Subscripts indicate whether the parameter relates to response time latencies (1) or within-word keystroke intervals (2). α = intercept; β = spelling H effect (for 2: mixture component 1); δ = spelling H effect (mixture component 2); θ = mixing proportion for mixture component 2; ϕ = autocorrelation; σ_s = error variance

Parameter	2.5%PI	$\hat{\mu}$	97.5%PI
α_1	910.89	1014.34	1126.01
α_2	92.95	109.24	128.48
β_1	-121.37	67.60	275.63
β_2	-3.81	14.51	36.08
δ_2	28.36	68.05	145.85
θ_2	0.17	0.31	0.49
ϕ_2	0.01	0.03	0.05
σ'_e	0.51	0.58	0.70
σ_e	0.24	0.27	0.30
σ_{diff}	0.12	0.16	0.20
σ	0.38	0.43	0.49

an autoregression model with a mixture component for the IKIs after typing onset. This indicates that the spelling H effect is to some extent probabilistic affecting the within-word keystrokes. The spelling effect H has a probability θ to affect the within-word keystroke intervals. This effect is indicated as δ . For the remaining $1 - \theta$ proportion of the within-word keystroke intervals spelling H has an effect β . This analysis renders that spelling H has no effect on the response time latencies but has a probability θ to affect the within-word keystroke intervals.