

Modelling disfluencies in copy-typing

Jens Roeser¹, Sven De Maeyer², Mark Torrance¹, Luuk Van Waes³, & Mariëlle Leijten³

¹ Department of Psychology, Nottingham Trent University, United Kingdom

² Faculty of Social Sciences, University of Antwerp, Belgium

³ Department of Management, University of Antwerp, Belgium

The analysis of keystroke data is typically performed on aggregates (mean, median). Aggregating data results in two basic problems. First, means are poor descriptives for non-normal distributed data. Second, using means artificially reduces information and adds bias to the estimates. In particular disfluencies affect these estimates. While many researchers use thresholds to define or trim data for certain purposes, disfluencies are important indicators for increased cognitive demands and are ought to be addressed for statistical inference. We implemented a series of Bayesian models and tested these models on a random sample of 100 participants (age 18-25) from the Dutch copy-task corpus. Comparing the predictive performance of these models showed strong support for models that account for disfluencies as a mixture process allowing for a principle method of detecting disfluencies during keyboard typing.

Keywords: Copy-task; keystroke modeling; autoregression; mixture models; Bayesian inference

Introduction

Writing research has made extensive use of keystroke-logging to capture typing process data. In particular disfluencies (loosly defined as relatively long intervals between subsequent keystrokes) are interesting as these are indicators of processing demands (Chukharev-Hudilainen, Sari-caoglu, Torrance, & Feng, 2019) which may arise on various levels of mental representation (Olive, 2014); for example, when preplanning syntactic dependencies (Roeser, Torrance, & Baguley, 2019), retrieving a lexical entry for a word or its spelling (Torrance, Rønneberg, Johansson, & Uppstad, 2016). At present there is no principled way of determining which keystroke intervals constitute a process disfluency. In this paper we present a series of statistical models that aimed at capturing the typing process and in particular process disfluencies.

Keystroke logs provide very rich information about the typing process. To reduce the complexity of the data, researchers have often performed analysis on the means, medians, standard deviations (SD) etc. of inter-keystroke intervals (the duration between two consecutive keystrokes), number

of pauses or pause durations, within-word keystroke intervals and many other variables (for an overview see Conijn, Roeser, & van Zaanen, 2019). Conijn et al. (2019) suggested that different aggregates are sensitive to processing difficulty that arises on different levels of mental representation. However, using data aggregates comes with a number of problems. First, a feature such as pause duration or number of pauses requires a definition what constitutes a pause. Researchs have often used a conventional but arbitrary threshold of 2 secs to define pauses (Alves & Limpo, 2015; Chukharev-Khudilaynen, 2014; Connelly, Dockrell, Walter, & Critten, 2012; Leijten & Van Waes, 2013). Second, aggregating data artificially reduces the information in the data requires trimming criteria (Hoaglin & Iglewicz, 1987) which remove certain data as outliers that might be valuable information about the writing process. For example, data aggregation disregards timecourse variation and autocorrelation between consecutive keystrokes (Eltahir, Salami, Ismail, & Lai, 2004). Third, parametric aggregates such the mean and the SD function assume that, for the aggregated data point to be representative for the sample, keystroke intervals must come from a normal distribution. This is not the case as keystroke intervals are zero-bound and therefore right-skewed.¹ Hence, aggregates are biased estimates of the typing process and therefore may lead to incorrect inference.

While the former problems are statistical problems they are

Correspondence concerning this article should be addressed to Jens Roeser, 50 Shakespeare St, Nottingham NG1 4FQ. E-mail: jens.roeser@ntu.ac.uk

¹In fact, the minimum size of keystroke intervals is determined by the time it takes to plan and execute the motor programm.

interlinked with a conceptual issue. Statistical models make assumptions about the nature of the data generating process. For example, most standard statistical models assume that the data come from a single process that generates normal distributed data. We already addressed that keystroke data are not normal distributed. Another problem is that the underlying mental process that generates keystroke data involves cognitive processing on different levels mental representation. Process difficulty on any of these levels inhibits the activation flow into lower levels of representation which then creates disfluencies (see Christiansen & Chater, 2016; Olive, 2014). For example, copying a sequence of letters involves visual encoding of the sequence of letters, buffering n-bigrams in memory, identifying the correct key(s) on the keyboard, and programming and executing the relevant motor codes. Difficulty at any of these levels of processing will inhibit processing lower levels of activation which results in typing disfluencies. Further, the duration of the delay might, to some extent, depend on the level of activation that is causing the lag. Thresholds are insufficient to distinguish fluent from disfluent typing to force the typing process into a single normal distribution. Crucially important information about the mental process would be removed. Planning the next keystroke may in some cases be more demanding, in particular when planning on phrase boundaries or due to the retrieval of the lexical form of a low frequent word. Also, long values are more likely for struggling writers. For incorrectly produced keystroke intervals, durations may or may not change depending on whether the error was noticed by the participant or create a dysfluency for a subsequent keystroke interval. Importantly, long keystrokes are not necessarily outliers and pauses are a central part of the writing process.

Modelling typing process data

As a guiding principle, we aim to produce statistical models that are representative for the theoretical process that creates the data at hand. All models aim to model the typing process based on unaggregated data, i.e. the lags between subsequent keypresses; e.g. $c^{\wedge}a^{\wedge}t$ for typing the word *cat*, where the caret \wedge indicates the difference between the time for pressing $\langle c \rangle$ and $\langle a \rangle$. Further, our models aim to provide a systematic way of addressing process disfluencies rather than requiring data removal or pausing thresholds. For example when the time for pressing $\langle a \rangle$ after $\langle c \rangle$ is unusually large, our models should make notice of this disfluency.

We implement statistical models that aim to capture these keystroke interval data. Statistical models allow us to characterise an underlying data generating process using parameters. For example if we assume that data come from a normal distribution, we can characterise this distribution with a value μ that indicates the centre of the distribution

and a scale σ^2 (i.e. the variance); this can be expressed as $y \sim Normal(\mu, \sigma^2)$. We will assume throughout that the inter-keystroke interval (IKI) data can be characterised with a log-normal distribution, because IKIs are zero-bound and therefore right-skewed. Linear mixed effects models (LMM), as in equation 1, are extensions of this logic which has been used widely for modelling keystroke data (Quené & Van den Bergh, 2004; Van Waes, Leijten, & Quinlan, 2010; Van Waes, Leijten, Roeser, Olive, & Grabowski, 2020). Crucially, the parameters of interest can be estimated while accounting for variance that is unknown and random. For example, some participants are faster typists than others. The variance associated with the i s participant, expressed as u_i , can be assumed to be normal distributed around 0 with a between participants variance σ_u^2 with $i = 1, \dots, I$, where I is the number of participants. After addressing individual typing-speed, the process that generates IKIs can be described by a mean μ and an error variance σ_e^2 . For model fit purposes, the mean μ was estimated using a non-centred parametrisation with weakly informative priors (Lambert, 2018). For the error variance we used a standard non-informative half-Cauchy prior (Gelman et al., 2014) as the variance must be larger than zero. Generally, we use weakly informative priors to aid model convergence. These priors help to reduce the parameter space to relatively plausible values. All models presented below share these properties. The following models differ as to how (1) timecourse differences and (2) disfluencies are addressed by these models.

A standard approach to model the variation between individual keystroke pairs (i.e. bigrams) is to assume that each bigram j with $j = 1, \dots, J$, where J is the total number of bigrams, is independent of the other bigrams and associated with a unknown variance w_j that is distributed around 0 with a between bigram variance σ_w^2 (see Van Waes, Leijten, Pauwaert, & Van Horenbeeck, 2019).

$$\begin{aligned}
 y_{ij} &\sim LogNormal(\mu + u_i + w_j, \sigma_e^2) \\
 &\text{where} \\
 \mu &= \mu_\alpha + \mu_\sigma * \mu_\eta \\
 \mu_\alpha &\sim Normal(5, 4) \\
 \mu_\sigma &\sim Normal(0, 10) \\
 \mu_\eta &\sim Normal(0, 1) \\
 u_i &\sim Normal(0, \sigma_u^2) \\
 w_j &\sim Normal(0, \sigma_w^2) \\
 \sigma_u, \sigma_w &\sim Normal(0, 2.5) \\
 \sigma_e &\sim Cauchy(0, 2.5) \\
 \text{constraint: } &\mu_\sigma, \sigma_u, \sigma_w, \sigma_e > 0
 \end{aligned} \tag{1}$$

Further, we can extend this model by assuming that larger variations in typing differences for bigrams depend on the

typing speed of each participant. For example, fast participants might show less variation between bigrams than slow participants. This assumption can be modelled by including by-participant slope adjustments for bigrams by introducing a variance-covariance matrix Σ_u ; LKJ prior with $\nu = 2.0$ (Lewandowski, Kurowicka, & Joe, 2009).

Typing as autoregressive process

The standard analysis can model variation that is associated with particular bigrams but does allow us to address disfluencies in the typing process beyond assuming they are subject to random statistical noise. Further, the standard analysis assumes that subsequent keystrokes are independent and thus exchangeable. If we assume that subsequent keystrokes are autocorrelated (Eltahir et al., 2004), a disfluency can be characterised as a slowdown relative to the previous keystroke interval. In other words, we can predict IKIs with the previous keystroke interval. This feature is at the heart of autoregressive models. These models capture the autocorrelation between subsequent keystrokes by introducing a parameter ϕ that predicts an IKI y_j on the basis of the IKI preceding it y_{j-1} ; see equation 2. ϕ_i was estimated separately for each participant with a pooled mean ϕ and error variance η^2 . Note, to allow comparisons between this and all other models we had exclude the first bigram for each participant, i.e. y_{ij+1} .

$$\begin{aligned}
 y_{ij} &\sim \text{LogNormal}(\mu + \phi_i * \log(y_{ij-1}) + u_i, \sigma_e^2) \\
 \text{where} \\
 \mu &= \mu_\alpha + \mu_\sigma * \mu_\eta \\
 \phi_i &\sim \text{Normal}(\phi, \eta^2) \\
 \mu_\alpha &\sim \text{Normal}(5, 4) \\
 \mu_\sigma &\sim \text{Normal}(0, 10) \\
 \mu_\eta, \phi &\sim \text{Normal}(0, 1) \\
 \eta &\sim \text{Cauchy}(0, 1) \\
 u_i &\sim \text{Normal}(0, \sigma_u^2) \\
 \sigma_u &\sim \text{Normal}(0, 2.5) \\
 \sigma_e &\sim \text{Cauchy}(0, 2.5) \\
 \text{constraint: } \mu_\sigma, \sigma_u, \sigma_e, \eta &> 0
 \end{aligned} \tag{2}$$

Typing as (autoregressive) finite mixture process

Writing data are generally speaking a combination of fluent typing and typing disfluencies. While disfluencies can be subject to processing on various levels of mental representations they will always be expressed as longer keystroke intervals. To capture this mixture of fluent and disfluent typing, we implemented a finite mixture model. Mixture models assume that data come from a combination of distributions.

For the present purpose we constrain the model to be finite. In other words, we fixed the number of underlying distributions to two, namely, 2 log-Gaussian (normal) distributions of which one represents fluent typing (shorter keystroke intervals) and the other process disfluencies (longer keystroke intervals). This model can be summarised as in equation 3, following Vasishth, Chopin, Ryder, and Nicenboim (2017). Both distributions have the same mean μ but for the first the parameter δ that was constrained to be positive was added. The mixing proportion θ_i , then, captures the probability of a data point to be part of either distribution (fluent or disfluent typing) for each participant i which a mean θ and variance τ^2 . The mixing proportion θ_i was transformed to range from 0 to 1 (inverse logit) where a value of 0 would indicate fluent typing and 1 indicates disfluency. As longer latencies are known to be associated with a larger variances for both response-time data in particular (Wagenmakers & Brown, 2007) and human motor behaviour in general (Schöner, 2002; Wing & Kristofferson, 1973), the variance σ_e^2 associated with the distribution of typing disfluencies was constrained to be larger than the variance for normal typing σ_e^2 .

$$\begin{aligned}
 y_{ij} &\sim \theta_i \cdot \text{LogNormal}(\mu + \delta + u_i + w_j, \sigma_{e'}^2) + \\
 &\quad (1 - \theta_i) \cdot \text{LogNormal}(\mu + u_i + w_j, \sigma_e^2) \\
 \text{where} \\
 \theta_i &= \text{InvLogit}(\theta_i) \\
 \mu &= \mu_\alpha + \mu_\sigma * \mu_\eta \\
 \sigma_{e'} &= \sigma + \sigma_{\text{diff}} \\
 \sigma_e &= \sigma - \sigma_{\text{diff}} \\
 \theta_i &\sim \text{Normal}(\theta, \tau^2) \\
 \mu_\alpha &\sim \text{Normal}(5, 4) \\
 \mu_\sigma &\sim \text{Normal}(0, 10) \\
 \theta, \delta, \mu_\eta, \sigma_{\text{diff}} &\sim \text{Normal}(0, 1) \\
 \tau &\sim \text{Cauchy}(0, 1) \\
 u_i &\sim \text{Normal}(0, \sigma_u^2) \\
 w_j &\sim \text{Normal}(0, \sigma_w^2) \\
 \sigma_u, \sigma_w &\sim \text{Normal}(0, 2.5) \\
 \sigma &\sim \text{Cauchy}(0, 2.5) \\
 \text{constraint: } \mu_\sigma, \sigma_u, \sigma_w, \delta, \tau, \sigma, \sigma_{\text{diff}}, \sigma_{e'}, \sigma_e &> 0
 \end{aligned} \tag{3}$$

Note that the former model implies that subsequent keystroke intervals are independent. This might be the case for disfluencies but fluent typing might involve autocorrelations. Therefore, we implemented another mixture model but replaced the bigram intercepts w_j in the previous model with an autoregressor $\phi_i * y_{ij-1}$ in the distribution that represents fluent typing, as in equation 2, but included random bigram intercepts for the distribution of disfluent typing intervals.

Method

We evaluate our models using data from a subset of the Dutch copy-task corpus (Leijten & Van Waes, 2013; Van Waes et al., 2019; Van Waes et al., 2020). The copy-task corpus consists of keystroke data collected in Inputlog, a Javascript-based web application, available on www.inputlog.net. In this analysis we focus on the consonant task because it is a demanding but short task that allows us to measure typing skills in a non-linguistic environment (Grabowski, Weinzierl, & Schmitt, 2010). The consonant task is relatively short but complex in the sense that participants have to look at the target string repeatedly to copy it accurately. In other words, for this task we need to be able to disentangle fluent and disfluent latencies. Participants saw and copy-typed four blocks of six consonant sequences “tjxgfl pgkfkq dtdrgt npwdvf” that do not occur adjacently in Dutch. The dependent variable are the inter-keystroke intervals (IKI) between two consecutive consonants (key presses). Spaces were excluded from the data. From this corpus we took a random sample of 100 participants (78 female, 22 male) in the age range between 18 and 25 years (median age = 22 years). We used decided this age range to keep the sample relatively homogeneous.

Results

Data overview

The data are visualized in Figure 1. The figure shows the variations of IKIs throughout the copy-typing process. Each grey line represents one participant and the coloured lines shown different measures of central tendency. The central tendency descriptives represent different properties of the distribution as can be seen in the density plot in the right margin. The mean is sensitive to larger IKIs and is therefore in the centre of the distribution while the mode and the median are closer to the most frequent value. All central tendencies vary depending on the position of the bigram in the string. The up-and-down pattern in the data suggests that IKIs are sensitive to the position of a bigram in a chunk rather than bigram identity or the position of the bigram in the entire sequence.

Model comparisons

All models were implemented as Bayesian models (see e.g. Gelman et al., 2014; Lambert, 2018; McElreath, 2016) in the probabilistic programming language Stan (Carpenter et al., 2016; Hoffman & Gelman, 2014; Stan Development Team, 2015a, 2015b). *R* and *Stan* code are available on GitHub for reproducibility. Models were fitted with weakly regulating priors and run with 30,000 iterations (15,000 warm-up) on 3 Markov chain Monte Carlo chains. Convergence

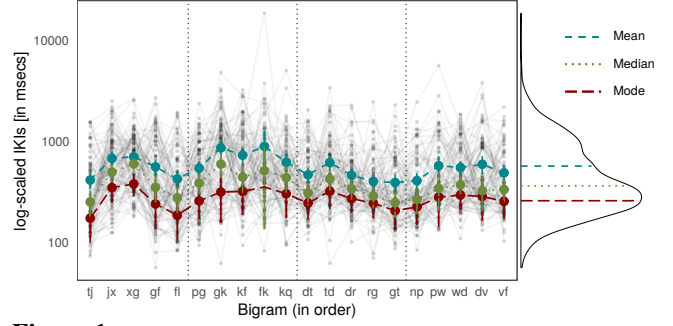


Figure 1

IKI data by participant over bigrams position shown in order. Shown are the raw data, each line representing one participant, and different measures of central tendency (with $2 \times \text{standard error [SE]}$). Vertical lines indicate the position of SPACES. The density distribution of IKI data is shown in the right margin.

was tested via the Rubin-Gelman statistic (Gelman & Rubin, 1992), traceplots and cross-validation (Vehtari, Gelman, & Gabry, 2015, 2017).

The predictive performance of the models was established using leave-one-out cross-validation which penalizes models with more parameters (see Farrell & Lewandowsky, 2018; Lambert, 2018; Lee & Wagenmakers, 2014; McElreath, 2016). The out-of-sample predictive performance was determined via Pareto smoothed importance-sampling (Vehtari et al., 2015, 2017) and estimated as sum of the expected log predictive density (\widehat{elpd}). \widehat{elpd} was used to compare the predictive quality of our models. Model comparisons can be found in Table 1. The best fitting model was revealed to be the mixture model M4 (see equation 3).

The second best performing model is the mixture model with the autoregressor *phi* for fluent typing. In other words, adding the autoregressor instead of random bigram intercepts for fluent typing did not improve the predictive performance of the model. In fact, the autoregressive model was found to be the model with the lowest predictive performance among the models tested. Modelling bigrams as random intercepts (with and without by-participant slope adjustments) was found to have a higher predictive performance compared to the autoregression model (Van Waes et al., 2010, 2020). Overall the best fit among the tested models was revealed by the mixture model with a distribution for fluent typing and a distribution for disfluencies while accounting for bigrams as random variance.

Model evaluation

The copy-typing process captured by the mixture model can be assessed using the posterior distributions of the param-

Table 1

Model comparisons $\Delta\widehat{elpd}$ expressed as expected log predictive density (\widehat{elpd}) with SE in parentheses. The bigrams column indicates how the models account for by-bigram variation. The top row shows the model with the highest predictive performance.

Model	Type	$\Delta\widehat{elpd}$	\widehat{elpd}	Bigrams
M4	MoG	0 (0)	-13857 (59)	Random intercepts
M5	AR + MoG	-14 (14)	-13871 (57)	Autoregressor ϕ and random intercepts
M2	LMM	-87 (24)	-13944 (53)	Random intercepts and slopes
M1	LMM	-153 (18)	-14010 (54)	Random intercepts
M3	AR	-223 (21)	-14080 (54)	Autoregressor ϕ

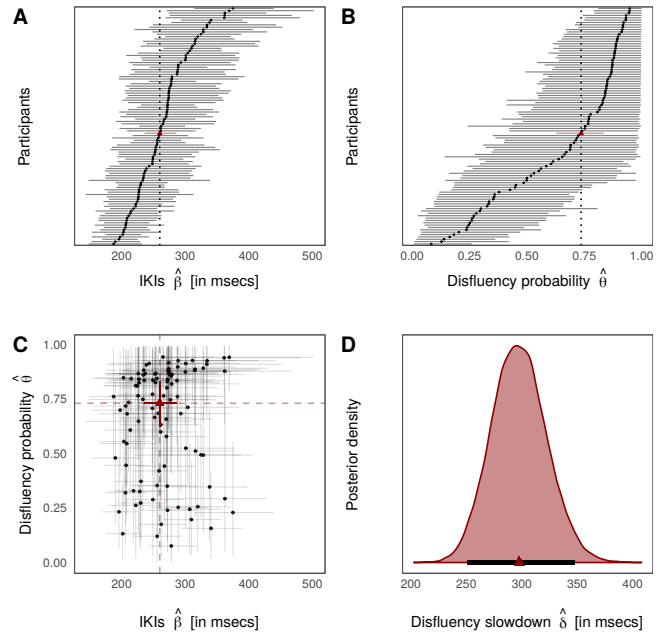
Note. LMM = linear mixed effects models; AR = Autoregressive model; MoG = Mixture of (log-)Gaussians

ter values. The process relevant model parameter values are illustrated in Figure 2. Firstly, IKIs, excluding disfluencies, are shown by participant in Figure 2A. The red line indicates the pooled overall parameter estimate for fluent typing $\hat{\beta}=259$ msec centered around a 95% PI of [235, 286]. For each participant the probability of disfluent typing is shown in Figure 2B. The overall disfluency probability (in red) was $\hat{\theta}=0.73$ centered around 95% PI[0.63, 0.83]. In other words, for the consonant task we observe 73% disfluent typing, and hence 27% fluent typing.

The y-axis in panel A and B are ordered by the average size of the respective values, thus the lines do not represent the same participants. In fact, Figure 2C suggests that the inferred latency for fluent typing and the probability to exhibit disfluencies are independent. Shown are the parameter estimates for each participant and the overall pooled estimates. In other words, the frequency of process disfluencies does not depend on the average typing speed. Fast as well as slow copy-typists can show low and high disfluency probabilities. Finally, the slowdown for disfluent typing is shown in Figure 2D. Disfluent typing in the consonant task is $\hat{\delta}=297$ msec (95% PI[251, 347]) slower than fluent typing.² Overall, there is a large probability of dysfluent typing in the consonant task. This makes it crucial to distinguish between keystroke intervals that represent fluent and disfluent typing. Mixture models can provide accurate estimates for fluent typing by accounting disfluencies in a principled matter; i.e. by modelling fluent and disfluent typing as a mixture process.

Discussion

Our aim was provide a statistical model of inter-keystroke intervals that addresses process disfluencies in a principled manner. We compared series of Bayesian models addressing this aim. Model comparisons showed that, among our models, process disfluencies can be captured best as a mixture process. This model captures disfluencies in the typing pro-

**Figure 2**

Posterior parameter values of the mixture model. Panel A shows by-participant IKIs and overall IKI value for fluent typing in red. Panel B shows by-participant disfluency probability (overall parameter value θ in red). Panel C shows fluent IKIs plotted against disfluency probability (red triangle indicates overall parameter value). Panel D shows the posterior distribution of the disfluency slowdown. All error bars are 95% probability intervals.

cess by assuming that typing data come from a combination

²The size of the disfluency δ could in principle vary by participant. For example, for faster participants might show larger disfluencies and slower participants might show smaller disfluencies. We addressed this possibility by allowing the disfluency parameter δ to vary for each participant. This model revealed a negligible gain over model M4 ($\Delta\widehat{elpd} = -3$, $SE=2$).

of two distributions, representing fluent and disfluent typing, of which each has a latent probability. This model allows us to extract reliable typing-interval estimates for fluent typing while accounting for process disfluencies in a principled manner.

This model provides a) overall and by-participant keystroke intervals for fluent typing, b) overall and by-participant disfluency probabilities, and c) an parameter estimate that captures the overall slowdown for disfluent trials. These parameter estimates are relevant on two levels. First, they allow us to characterise the writing task at hand. For example, we observe that copy-typing non-lexical strings of consonants is ridden with disfluencies. Second, by-participant parameter estimates allow us to extract typing characteristics for individual typists. In particular, we extracted each participant's fluent typing speed and the probability of disfluencies exhibited by each participant. Taken both, overall and individual parameter estimates together, we can determine whether an individual typist was a fast / slow typist or had unusually high / low probability to exhibit disfluencies.

The central advantage of using mixture models to account for typing disfluencies is that we can by-pass the use of threshold values to define disfluencies. From the raw data it is not possible to know which data are disfluencies a priori. Using threshold values merely ignores that some participants are generally lower typists and some tasks are easier or more difficult. Mixture models allow us to capture disfluencies as a latent process in a principled way. This is important because the mixture model presented in this paper takes into account that a disfluency is relative to an individual's typing speed. Therefore, this would allow to test predictions about typing disfluencies in certain populations such as very young or old typists, language learners or individuals with genuine typing difficulty after account for individual differences in typing speed. Further while we used a relatively difficult copy-task to test our models, other writing tasks are bound to lead to other disfluency estimates. In fact, the presented model can be used to test hypothesis about psychological factors that might lead to varying proportions of disfluencies, and therefore changes in cognitive demands, in the writing process. As an avenue for future research, mixture models as presented in this paper can be used for different types of writing tasks and particular populations.

Writing involves processing on various cognitive levels, from generating an idea, retrieving lexical entries to the encoding of spelling. As activation between these levels cascades from higher to lower levels of representation, a delay at any of these levels will cause disfluencies. In other words, disfluencies are not just an interesting but a central part of the writing process. Mixture models allow us to address the assumed combination of processes that underlies writing by mapping these of a combination of distributions. Using a combination

of two distributions is merely reducing this combination into normal, fluent, typing and typing that involves processing delays on higher levels (i.e. disfluencies). If one assumes that the size of the disfluency depends on the inhibited upstream process, one may implement this as an additional mixture component (see e.g. Baaijen, Galbraith, & Glopper, 2012) for address different types of disfluencies (Wengelin, 2001).

References

- Alves, R. A., & Limpo, T. (2015). Progress in written language bursts, pauses, transcription, and written composition across schooling. *Scientific Studies of Reading*, 19(5), 374–391.
- Baaijen, V. M., Galbraith, D., & Glopper, K. de. (2012). Keystroke analysis: Reflections on procedures and measures. *Written Communication*, 29(3), 246–277.
- Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., ... Riddell, A. (2016). Stan: A probabilistic programming language. *Journal of Statistical Software*, 20.
- Christiansen, M. H., & Chater, N. (2016). The now-or-never bottleneck: A fundamental constraint on language. *Behavioral and Brain Sciences*, 39, 1–72. <https://doi.org/http://dx.doi.org/10.1017/S0140525X1500031X>
- Chukharev-Hudilainen, E., Saricaoglu, A., Torrance, M., & Feng, H.-H. (2019). Combined deployable keystroke logging and eyetracking for investigating L2 writing fluency. *Studies in Second Language Acquisition*, 41(3), 583–604.
- Chukharev-Khudilaynen, E. (2014). Pauses in spontaneous written communication: A keystroke logging study. *Journal of Writing Research*, 6(1), 61–84.
- Conijn, R., Roeser, J., & van Zaanen, M. (2019). Understanding the keystroke log: The effect of writing task on keystroke features. *Reading and Writing*, 32(9), 2353–2374.
- Connelly, V., Dockrell, J. E., Walter, K., & Critten, S. (2012). Predicting the quality of composition and written language bursts from oral language, spelling, and handwriting skills in children with and without specific language impairment. *Written Communication*, 29(3), 278–302.
- Eltahir, W. E., Salami, M., Ismail, A. F., & Lai, W. (2004). Dynamic keystroke analysis using AR model. In *IEEE international conference on industrial technology* (Vol. 3, pp. 1555–1560). IEEE.
- Farrell, S., & Lewandowsky, S. (2018). *Computational modeling of cognition and behavior*. Cambridge University Press.

- Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2014). *Bayesian data analysis* (3rd ed.). Chapman; Hall/CRC.
- Gelman, A., & Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science*, 7(4), 457–472.
- Grabowski, J., Weinzierl, C., & Schmitt, M. (2010). Second and fourth graders' copying ability: From graphical to linguistic processing. *Journal of Research in Reading*, 33(1), 39–53.
- Hoaglin, D. C., & Iglewicz, B. (1987). Fine-tuning some resistant rules for outlier labeling. *Journal of the American Statistical Association*, 82(400), 1147–1149.
- Hoffman, M. D., & Gelman, A. (2014). The No-U-Turn sampler: Adaptively setting path lengths in Hamiltonian Monte Carlo. *Journal of Machine Learning Research*, 15(1), 1593–1623.
- Lambert, B. (2018). *A student's guide to Bayesian statistics*. Sage.
- Lee, M. D., & Wagenmakers, E.-J. (2014). *Bayesian cognitive modeling: A practical course*. Cambridge University Press.
- Leijten, M., & Van Waes, L. (2013). Keystroke logging in writing research: Using Inputlog to analyze and visualize writing processes. *Written Communication*, 30(3), 358–392.
- Lewandowski, D., Kurowicka, D., & Joe, H. (2009). Generating random correlation matrices based on vines and extended onion method. *Journal of Multivariate Analysis*, 100(9), 1989–2001.
- McElreath, R. (2016). *Statistical rethinking: A bayesian course with examples in R and Stan*. CRC Press.
- Olive, T. (2014). Toward a parallel and cascading model of the writing system: A review of research on writing processes coordination. *Journal of Writing Research*, 6(2), 173–194.
- Quené, H., & Van den Bergh, H. (2004). On multi-level modeling of data from repeated measures designs: A tutorial. *Speech Communication*, 43(1-2), 103–121.
- Roeser, J., Torrance, M., & Baguley, T. (2019). Advance planning in written and spoken sentence production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 45(11), 1983–2009.
- Schöner, G. (2002). Timing, clocks, and dynamical systems. *Brain and Cognition*, 48(1), 31–51.
- Stan Development Team. (2015a). Stan: A C++ library for probability and sampling. <http://mc-stan.org/>.
- Stan Development Team. (2015b). Stan modeling language user's guide and reference manual. <http://mc-stan.org/>.
- Torrance, M., Rønneberg, V., Johansson, C., & Uppstad, P. H. (2016). Adolescent weak decoders writing in a shallow orthography: Process and product. *Scientific Studies of Reading*, 20(5), 375–388.
- Van Waes, L., Leijten, M., Pauwaert, T., & Van Horenbeeck, E. (2019). A multilingual copy task: Measuring typing and motor skills in writing with inputlog. *Journal of Open Research Software*, 7(30), 1–8.
- Van Waes, L., Leijten, M., & Quinlan, T. (2010). Reading during sentence composing and error correction: A multi-level analysis of the influences of task complexity. *Reading and Writing*, 23(7), 803–834. <https://doi.org/10.1007/s11145-009-9190-x>
- Van Waes, L., Leijten, M., Roeser, J., Olive, T., & Grabowski, J. (2020). Designing a copy task to measure typing and motor skills in writing research. *Journal of Writing Research*.
- Vasishth, S., Chopin, N., Ryder, R., & Nicenboim, B. (2017). Modelling dependency completion in sentence comprehension as a Bayesian hierarchical mixture process: A case study involving Chinese relative clauses. *ArXiv E-Prints*.
- Vehtari, A., Gelman, A., & Gabry, J. (2015). Pareto smoothed importance sampling. *arXiv Preprint arXiv:1507.02646*.
- Vehtari, A., Gelman, A., & Gabry, J. (2017). Practical bayesian model evaluation using leave-one-out cross-validation and WAIC. *Statistics and Computing*, 27(5), 1413–1432.
- Wagenmakers, E.-J., & Brown, S. (2007). On the linear relation between the mean and the standard deviation of a response time distribution. *Psychological Review*, 114(3), 830–841. <https://doi.org/10.1037/0033-295X.114.3.830>
- Wengelin, Å. (2001). Disfluencies in writing – Are they like in speaking? In *ISCA tutorial and research workshop (ITRW) on disfluency in spontaneous speech*.
- Wing, A. M., & Kristofferson, A. B. (1973). Response delays and the timing of discrete motor responses. *Perception & Psychophysics*, 14(1), 5–12.