

Modelling disfluencies in copy-typing as mixture process

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

Author Note

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

Abstract

The analysis of keystroke timing data typically involves the calculation of summary statistics such as the mean inter-keystroke interval, pause frequencies etc. There are two fundamental problems with this: first, parametric descriptives for non-normal distributed data are going to be biased estimates and neglect information associated with variance; second, pauses and pause-related process measures are defined using threshold value which are, in principle, arbitrary. We implemented a series of Bayesian models that aimed to address both issues by (a) providing reliable typing estimates and (b) statistically detecting process disfluencies. We tested these models on a random sample of 100 participants from the Dutch copy-task corpus. Our results illustrate how typing disfluencies can be statistically determined as a finite mixture process. The typing task was characterized by this mixture model as a combination of fluent and disfluent typing associated with parameters that capture the magnitude and the probability of disfluencies. Mixture models provide a principle method to detect disfluencies in keyboard typing data.

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

Modelling disfluencies in copy-typing as mixture process

Introduction

Writing research has made extensive use of keystroke-logging to capture typing process data. In particular process disfluencies (loosely defined as relatively long intervals between subsequent keystrokes) are interesting to develop an understanding of the individuals writing progress. This is because language production is typically thought of as a cascade from the mental generation of a message, into grammatical processing and finally the generation and execution of motor codes that serve the transition of an idea. This can be found in theoretical models of speech (Bock & Ferreira, 2014), handwriting (Van Galen, 1991) and keyboard typing (Hayes, 2012). Disfluencies at the execution stage are therefore indicators of processing demands that arise on higher levels of mental representation (Christiansen & Chater, 2016; Olive, 2014); for example, when preplanning syntactic dependencies (Roeser, Torrance, & Baguley, 2019) or retrieving a lexical entry for a word or its spelling (Torrance, Rønneberg, Johansson, & Uppstad, 2016). At present there is no principled way of detecting keystroke lags that 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 rich information about the typing process. From this log, researchers can calculate different process measures including measures of writing fluency (Chukharev-Hudilainen, Saricaoglu, Torrance, & Feng, 2019). To name a few, researchers have performed data analysis on means, medians, standard deviations (SD) etc. of inter-keystroke intervals (the duration between two consecutive keystrokes), number of pauses or pause duration, within-word keystroke intervals and many other variables (for an overview see Conijn, Roeser, & van Zaanen, 2019). Conijn et al. (2019) suggested that these aggregates are sensitive to processing difficulty that arises on different levels of mental representation. However, there are two substantial problems tight to this.

First, pause frequency, bursts and related measures (Alves & Limpo, 2015) require a definition of what passes as a pause (Wengelin, 2006), i.e. a pause criterion often set to 2 secs (Chanquoy, Foulin, & Fayol, 1996; Kaufer, Hayes, & Flower, 1986; Sullivan & Lindgren, 2002; Wengelin, 2002) or some other lower bound (Chukharev-Khudilaynen, 2014; Connelly, Dockrell, Walter, & Critten, 2012; Leijten & Van Waes, 2013). Researchers have stipulated pause thresholds specific to their research purposes and based on prior research. However, ideally, this threshold would need to be specific to both the writing task and the skills of the typist (Wengelin, 2006). For example, when comparing the frequency of pauses larger than 2 secs for a dyslexic and a normal typist one might observe more pauses for the dyslexic because a) 2 secs are indeed not unusual transitions between two keystrokes for a dyslexic or b) because pauses for the normal typist are typically shorter than 2 secs and therefore unobserved given the 2 secs pause criterion (Wengelin, 2001).

Second, data aggregation results in the loss of important information about disfluencies and time course variation. Even if this variation is not of interest to answer a research question, parametric aggregates such the mean and the SD are biased estimates of the typing process. This is because parametric aggregates assume, by definition, that the data must come from a normal distribution for the summary statistic to be representative for the sample. This is not the case as keystroke intervals are zero-bound and therefore right-skewed.¹ Therefore, data aggregation may lead to incorrect inference based on biased parameter values. To prevent biased parameter values, i.e. to ensure normal distributed data, summary statistics involve data trimming (Hoaglin & Iglewicz, 1987) to remove data that were *a priori* considered outliers. However, the removal of such disfluencies affects groups of struggling writers more than normal writers and therefore skews the information in the data.

A central methodological challenge with implications for writing research (Hayes,

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

2012; Kaufer et al., 1986; Wengelin, 2006) is the detection of writing disfluencies. We addressed this problem by implementing statistical models that aim to capture the nature of the data generating process (i.e. keyboard typing). Crucially we want these models to provide reliable estimates to an individuals typing performance without subjecting the data to trimming criteria, or threshold values and aggregation.

Modelling typing process data

As a guiding principle, we aim to produce statistical models that are in line with the mental process that creates the data. The typing process data are measures of the lag between subsequent keypresses (the inter-keystroke interval; IKI), for example, the transitions $c^{\wedge}a^{\wedge}t$ for the word *cat* where \wedge indicates the typing difference between pressing $\langle c \rangle$ and $\langle a \rangle$. Further, our models should provide a systematic way of addressing process disfluencies; when the lag before pressing $\langle a \rangle$ is unusually large. We implemented a series of possible models for the keystroke data. The predictive performance of these models will be compared in the Results section.

Statistical models can be used to characterize an underlying data generating process as a function with parameter values. For example, if we assume that the process that we are interested in can be described as normal distribution, we need a mean μ and a variance σ^2 to draw this distribution. The values of the parameters μ and σ^2 are unknown and often expected to vary across task demands and population. This model can be written as $y \sim Normal(\mu, \sigma^2)$; the data y come from a process that follows a normal distribution with an unknown mean μ and an unknown error variance σ^2 . Our statistical models need to determine values for these parameters that can be considered reliable. Bayesian models, as used in this paper, are ideal for reliable parameter estimation as they allow us to derive a probability distribution of the parameter value of interest (Farrell & Lewandowsky, 2018; Gelman et al., 2014; Lee & Wagenmakers, 2014). To achieve this, Bayesian models require the explicit inclusion of prior information about the distribution of the parameters of

interest. For small data sets even vague prior information influences the posterior (inferred parameter estimates) but for larger data sets the posterior is overcome by the data (i.e. automatic Ockam’s razor; Jefferys & Berger, 1992). In the present paper, priors are used to aid model convergence by constraining the parameter space (i.e. using weakly regulating priors; Lambert, 2018; McElreath, 2016).

We assume throughout that IKIs can be characterized as log-normal distributed because IKIs are zero-bound (Baayen, 2008), as in equation 1.

$$y_{ij} \sim \text{LogNormal}(\mu + u_i + w_j, \sigma_e^2) \quad (1)$$

Equation 1 is an extension of the simple example above, which is generally referred to as linear mixed effects models (LMM) which has been used to model keystroke data (Quené & Van den Bergh, 2004; Van Waes, Leijten, & Quinlan, 2010; Van Waes, Leijten, Roeser, Olive, & Grabowski, 2020). The advantage of LMMs is that the parameters of interest, the mean μ , can be estimated while accounting for assumed sources of random error variance. For example, some participants are faster typists than others. These differences 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 (see 2).

$$\begin{aligned} u_i &\sim \text{Normal}(0, \sigma_u^2) \\ \sigma_u &\sim \text{Normal}(0, 2.5) \\ \text{constraint: } \sigma_u &> 0 \end{aligned} \quad (2)$$

A standard approach to model the variation between individual keystroke pairs (i.e. bigrams), as shown before in equation 1, 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 as in equation 3 (see Van Waes, Leijten, Pauwaert, & Van Horenbeeck, 2019).

$$\begin{aligned}
 w_j &\sim \text{Normal}(0, \sigma_w^2) \\
 \sigma_w &\sim \text{Normal}(0, 2.5) \\
 \text{constraint: } &\sigma_w > 0
 \end{aligned} \tag{3}$$

For model fit purposes, we parameterised the mean μ as non-centred with weakly regulating priors (Lambert, 2018) as shown in equation 4.

$$\begin{aligned}
 \mu &= \mu_\alpha + \mu_\sigma * \mu_\eta \\
 \mu_\alpha &\sim \text{Normal}(5, 2) \\
 \mu_\sigma &\sim \text{Normal}(0, 10) \\
 \mu_\eta &\sim \text{Normal}(0, 1) \\
 \text{constraint: } &\mu_\sigma > 0
 \end{aligned} \tag{4}$$

For the unexplained variance σ_e^2 we used a uninformative half-Cauchy prior (equation 5; Gelman et al., 2014).

$$\begin{aligned}
 \sigma_e &\sim \text{Cauchy}(0, 2.5) \\
 \text{constraint: } &\sigma_e > 0
 \end{aligned} \tag{5}$$

Further, we can extent 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 previous model captures variation associated with particular bigrams but assumes that disfluencies are subject to random noise. Further, the standard analysis assumes that subsequent keystrokes are independent and thus exchangeable. IKIs are not necessarily independent but IKI_i might be related to IKI_{i-1} preceding it. In other words, we can predict an IKI with the previous keystroke and captured their relationship with a parameter ϕ ; see equation 6. This is called an autoregressive process (Eltahir, Salami, Ismail, & Lai, 2004). This model captures disfluencies as slowdown relative to a previous keystroke. The autocorrelation was assumed to vary for each participant ϕ_i 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 y_{i1} for each participant.

$$\begin{aligned}
 y_{ij} &\sim \text{LogNormal}(\mu + \phi_i * \log(y_{ij-1}) + u_i, \sigma_e^2) \\
 &\text{where} \\
 \phi_i &\sim \text{Normal}(\phi, \eta^2) \\
 \phi &\sim \text{Normal}(0, 1) \\
 \eta &\sim \text{Cauchy}(0, 1) \\
 &\text{constraint: } \eta > 0
 \end{aligned} \tag{6}$$

Typing as (autoregressive) mixture process

Disfluencies during the typing process can be captured in finite mixture models. 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 tying (shorter IKIs) and the other represents disfluencies (longer IKIs). This model can be summarised as in equation 7, following Vasishth, Chopin, Ryder,

and Nicenboim (2017). The first and second line are the sum of two log-normal distributions of which the first distribution has a mixing proportion (weight) θ and the other distribution receives the remaining proportion $1 - \theta$. Both distributions have the same mean μ but the parameter δ that added to the first distribution and constrained to be positive. Thus, δ captures the magnitude of the disfluency. The mixing proportion θ_i , then, captures the probability of disfluent IKIs for each participant i .

$$\begin{aligned}
 y_{ij} &\sim \theta_i \cdot \text{LogNormal}(\mu + \delta + u_i + w_j, \sigma_e^2) + \\
 &\quad (1 - \theta_i) \cdot \text{LogNormal}(\mu + u_i + w_j, \sigma_e^2) \\
 &\quad \text{where} \\
 &\quad \delta \sim \text{Normal}(0, 1) \\
 &\quad \text{constraint: } \delta > 0
 \end{aligned} \tag{7}$$

The hyper-parameter θ captures the population disfluency probability (with an error variance τ^2) as shown in equation 8. 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.

$$\begin{aligned}
 \theta_i &= \text{InvertedLogit}(\theta_i) \\
 \theta_i &\sim \text{Normal}(\theta, \tau^2) \\
 \theta &\sim \text{Normal}(0, 1) \\
 \tau &\sim \text{Cauchy}(0, 1) \\
 &\text{constraint: } \tau > 0
 \end{aligned} \tag{8}$$

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 as shown in 9 (see Vasishth et al., 2017; Vasishth, Jäger, & Nicenboim, 2017).

$$\begin{aligned}
 \sigma_{e'} &= \sigma + \sigma_{\text{diff}} \\
 \sigma_e &= \sigma - \sigma_{\text{diff}} \\
 \sigma_{\text{diff}} &\sim \text{Normal}(0, 1) \\
 \sigma &\sim \text{Cauchy}(0, 2.5) \\
 \text{constraint: } &\sigma, \sigma_{\text{diff}}, \sigma_{e'}, \sigma_e > 0
 \end{aligned}
 \tag{9}$$

Note that the mixture model, as well as the LMM, implies that subsequent keystroke intervals are independent. This might be the case for disfluencies but subsequent IKIs in fluent typing might involve autocorrelations. Therefore, we implemented another mixture model but replaced the bigram intercepts w_j , in the distribution that represents fluent typing in equation 7, with an autoregressor $\phi_i * y_{ij-1}$, as in equation 6; random bigram intercepts were kept for the distribution of disfluent typing intervals.

Method

To test which model captures the typing process best, we applied a series of models as described in the previous section to data from a subset of the Dutch copy-task corpus (Leijten & Van Waes, 2013; Van Waes et al., 2019; Van Waes et al., 2020). An overview of all models can be found in Table 1.

The copy-task corpus consists of keystroke data collected in Inputlog, a Javascript-based web application available on www.inputlog.net. In a set of different contexts participants have to produce keyboard typed responses. In this analysis we focus on the consonant task. Participants saw and copy-typed four blocks of six consonant sequences “tjxgfl pgkfkq dtdrgt npwdvf” that do not occur adjacently in Dutch. This task allows us to measure typing skills in a non-linguistic environment (Grabowski, Weinzierl, &

Table 1

Overview of typing process models. All models were fitted with random intercepts for participants.

Models	Type	Description
M1	LMM	Random intercepts for bigram order
M2	LMM	As M1 but with by-participant slope adjustments for bigram order
M3	AR	ϕ for autocorrelation between subsequent IKIs
M4	MoG	Mixture of normal typing and disfluencies δ with probability θ
M5	AR + MoG	As M4 but autocorrelation ϕ , as in M3, for normal typing

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

Schmitt, 2010). Importantly for the present purpose, the task requires repeated looks at the target string to copy it accurately which is relative to the participants memory span. This results in a combination of fluent typing and typing interruptions. In other words, for this task we need to be able to disentangle fluent and disfluent IKIs. We used a random sample of 100 participants (78 females, 22 males) in the age range between 18 and 25 years (median age = 22 years) to keep the sample relatively homogeneous.

Results

Data overview

The raw data are visualized in Figure 1. Each grey line represents one participant and the coloured lines shown different measures of central tendency. This figure highlights that the consonants task is a combination of fluent and disfluent typing. This combination

is expressed in the up-and-down pattern but also in the bimodality of the density function in the right margin. Further using central tendency descriptives highlight different properties of the distribution. The mean is sensitive to larger IKIs and is therefore not at the peak of the distribution while the mode and the median are relative close to the peak. All central tendencies vary depending on the position of the bigram in the string.

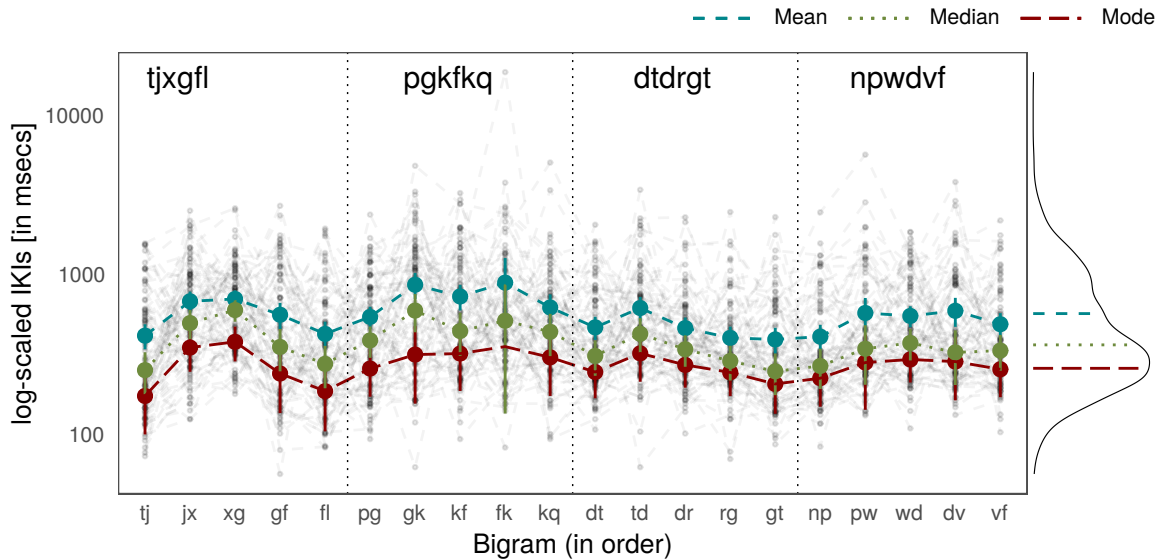


Figure 1

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 $2 \times$ standard error [SE]). Vertical lines indicate the position of SPACES. The density distribution of IKI data is shown in the right margin.

Model fit

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 (github.com/jensroes/Typing-disfluency). Models were fitted with 30,000 iterations (15,000 warm-up) on 3 Markov chain Monte

Carlo chains. Convergence 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 2. The mixture model M4 (see equation 7) revealed the highest predictive performance.

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. 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.

Parameter evaluation

The copy-typing process captured by the mixture model can be characterized with the posterior distributions of the model’s parameter values. The process relevant parameters 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 msecs centred 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 centred around 95% PI[0.63, 0.83]. In other words, for the consonant task

Table 2

Model comparisons expressed as expected log predictive density (\widehat{elpd}). The top row shows the model with the highest predictive performance. Differences in predictive performance are shown as $\Delta\widehat{elpd}$. Standard errors (SE) are shown in brackets.

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

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

we observe 73% disfluent typing and 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, 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 $\hat{\delta}=297$ msecs (95% PI[251, 347]) slower than

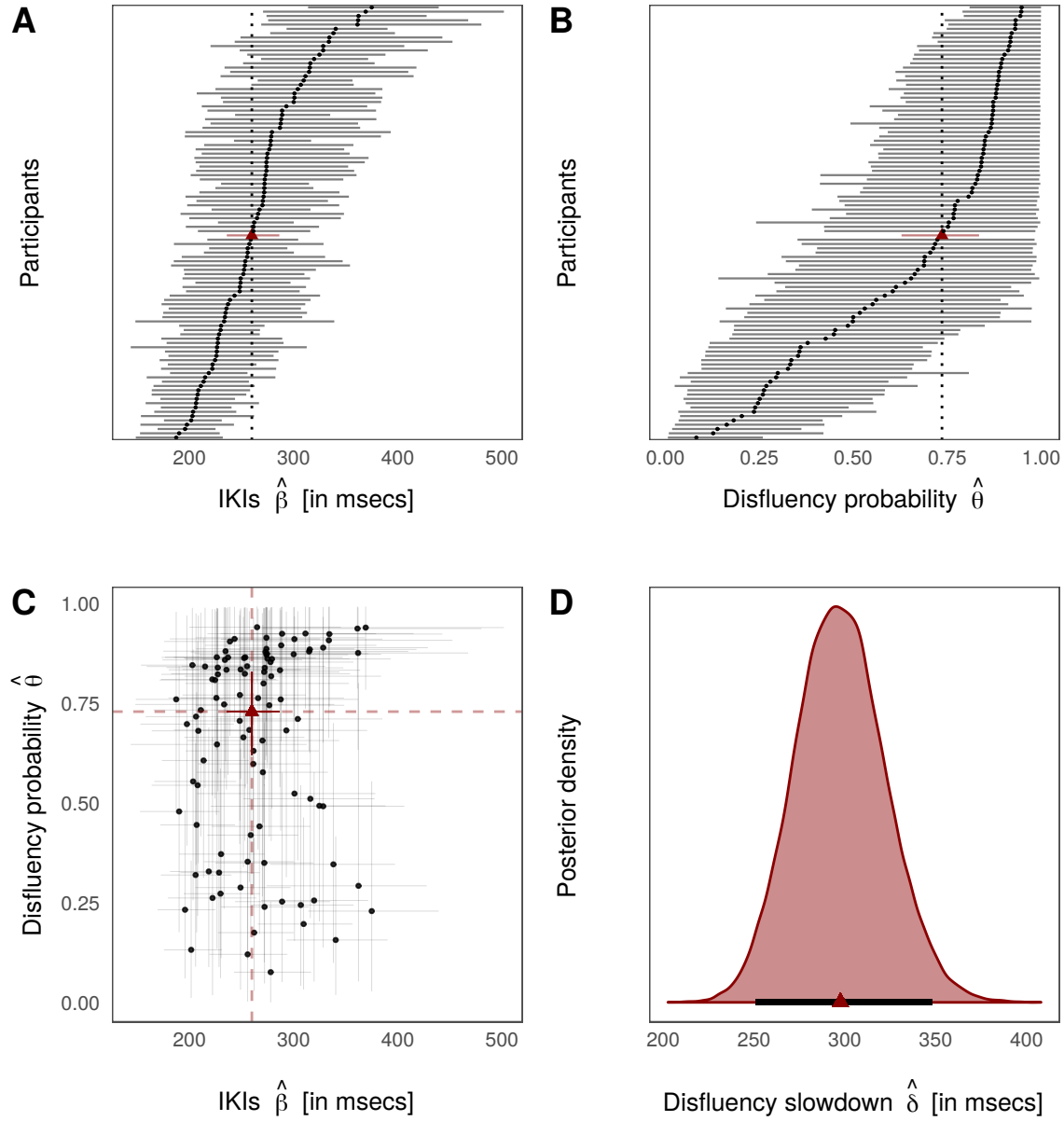
fluent typing.² Overall, for the consonant task it is crucial to distinguish between fluent and disfluent typing because disfluencies are indeed more common. In other words, the task demands do not inhibit keystroke intervals throughout the task but a large proportion of produced bigrams. Mixture models can provide accurate estimates for fluent typing while account for disfluencies by modelling fluent and disfluent typing as a mixture process.

Discussion

Our aim was to provide a statistical model of inter-keystroke intervals that addresses process disfluencies in a principled manner. We compared a series of Bayesian models addressing this aim. Model comparisons showed that process disfluencies can be captured as a mixture process for the consonant copy-task. This model allows us to extract reliable typing-interval estimates for fluent typing while accounting for process disfluencies by modelling fluent and disfluent typing as a combination of two distributions with a latent mixing ratio.

This model provides a probability distribution of the parameter values for a) fluent typing, overall and by-participant, b) the disfluency probability, overall and by-participant, and c) the disfluency magnitude (i.e. typing slowdown). These parameter estimates are relevant on two levels. First, they allows us to characterize the writing task at hand. For example, we observed that copy-typing non-lexical strings of consonants shows indeed a larger proportion of disfluent compared to fluent typing. Second, by-participant parameter estimates allow us to extract typing characteristics for individual typists. In particular, we extracted each participants fluent typing speed and the probability of disfluencies exhibited by each participant. The probability of disfluencies in this task can be understood as

² Faster participants might show larger disfluencies magnitudes; i.e. the size of the disfluency δ may vary by participant. This was not supported for the consonant copy-task. Allowing δ to vary by participant renders a negligible gain over model M4 ($\Delta\widehat{elpd}=3$, SE=2); holding θ constant while allowing δ to vary by participant revealed a lower predictive performance ($\Delta\widehat{elpd}=-48$, SE=14).

**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.

measures of memory span (Grabowski et al., 2010). If participants with a smaller memory span look more often to the target string, they will show a larger proportion of typing disfluencies. Taken together with the overall and individual parameter estimates, we can determine whether an individual was a fast / slow typist or had unusually high / low probability to exhibit disfluencies compared to the population estimates. Thus, the model can be used diagnostically to identify participants with larger disfluency probabilities or to compare pausing across groups of participants.

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. Using threshold values ignores that some participants are generally slower typists and some tasks are more difficult. Mixture models allow us to capture disfluencies as a latent process in a principled way. This is important because our mixture models take into account that a disfluency is relative to an individuals' typing speed and the task at hand (Wengelin, 2006). Therefore, these models allow us to test predictions about typing disfluencies in certain population such as learning typists, L2 typists and individuals with genuine typing difficulty after account for individual differences in typing speed or vice versa. Further while we used copy-task data, other writing tasks will lead to other disfluency estimates. In other words, the presented model can be used to test hypothesis about psychological factors that might affect the ratio of disfluencies 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 levels of mental representation. As activation cascades from higher to lower levels of representation, a delay on any of these levels causes disfluencies. While we distinguished fluent and disfluent typing in a binary way, processing difficulty on different levels might be associated with different disfluency magnitudes and might be cumulative. If the size of the disfluency is assumed to depend on the inhibited

process upstream or combination of processes, this can be implemented as additional mixture component(s) (Baaijen, Galbraith, & Glopper, 2012) to address different types of disfluencies (Wengelin, 2001). In other words extensions of mixture models allow us to test different hypotheses about the cascade of processes involved in writing and language production.

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.
- Baayen, R. H. (2008). *Analyzing linguistic data. A practical introduction to statistics using R*. Cambridge: Cambridge University Press.
- Bock, J. K., & Ferreira, V. S. (2014). Syntactically speaking. In M. Goldrick, V. S. Ferreira, & M. Miozzo (Eds.), *The Oxford Handbook of Language Production* (pp. 21–46). Oxford: Oxford University Press.
- 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.
- Chanquoy, L., Foulin, J.-N., & Fayol, M. (1996). Writing in adults: A real-time approach. In G. Rijlaarsdam, H. Van den Bergh, & M. Couzijn (Eds.), *Theories, models and methodology in writing research* (pp. 36–44). Amsterdam: Amsterdam University Press.
- 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.

Hayes, J. R. (2012). Evidence from language bursts, revision, and transcription for

translation and its relation to other writing processes. In M. Fayol, D. Alamargot, & V. Berninger (Eds.), *Translation of thought to written text while composing* (pp. 15–25). New York, NY: Psychology Press.

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.

Jefferys, W. H., & Berger, J. O. (1992). Ockham's razor and Bayesian analysis. *American Scientist*, 80(1), 64–72.

Kaufer, D. S., Hayes, J. R., & Flower, L. (1986). Composing written sentences. *Research in the Teaching of English*, 20(2), 121–140.

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/>.

Sullivan, K. P. H., & Lindgren, E. (2002). Self-assessment in autonomous computer-aided second language writing. *ELT Journal*, 56(3), 258–266.

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 Galen, G. P. (1991). Handwriting: Issues for a psychomotor theory. *Human Movement Science*, 10(2), 165–191.

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 multilevel 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*.

Vasishth, S., Jäger, L. A., & Nicenboim, B. (2017). Feature overwriting as a finite mixture process: Evidence from comprehension data. *arXiv Preprint arXiv:1703.04081*.

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*.

Wengelin, Å. (2002). *Text production in adults with reading and writing difficulties* (PhD thesis). Göteborg University.

Wengelin, Å. (2006). Examining pauses in writing: Theory, methods and empirical data. In K. P. H. Sullivan & E. Lindgren (Eds.), *Computer keystroke logging and writing: Methods and applications* (Vol. 18, pp. 107–130). Amsterdam: Elsevier.

Wing, A. M., & Kristofferson, A. B. (1973). Response delays and the timing of discrete motor responses. *Perception & Psychophysics*, 14(1), 5–12.