Representation of Geminate Letters in Typed Word Production

An Honors Thesis for the Department of Psychology

Renee Brody

Tufts University, 2019

I would like to thank Dr. Uriel Cohen Priva for supplying access to the raw data from his 2010 research on constructing typing-time corpora.

Abstract

Writing a word requires its spelling information to be retrieved as a set of abstract grapheme identities, and then for those graphemes to be held in working memory until articulation is finished. Geminate, or doubled, letters are not represented as independent copies, but rather are processed as a single unit on at least some level of the writing process. The present study compared typed production of words with and without geminates in a partial replication of research by Kandel et al. (2013, 2014). Forty-two participants completed a spelling-to-dictation task, typing each target word three times in immediate succession. A typing-time corpus was constructed from the data. Mixed-effects models were used to analyze latencies and IKIs compared to expected values from the corpus. Geminate words had an inhibitory effect on word latencies, and a facilitatory effect on the IKIs of the 2nd and 3rd productions of targets. Effects of geminates on IKIs of the initial production were not significant, suggesting that the facilitatory and inhibitory effects canceled each other out during the initial production. The results are consistent with dual-loop processing models, but are not consistent with Kandel et al.'s findings of cascading effects.

Representation of Geminate Letters in Typed Word Production

Written language is in many ways as central to everyday life as spoken language, so much so that we take for granted how rapidly and automatically we mentally convert sounds into strings of letters and back again. The representation of geminate, or doubled, letters in a written word is one of many phenomena not yet fully explained by our models of writing and reading. Geminate letters (e.g. ***bb*** in *ra**bb**it*) have been shown to be processed as a singular doubled grapheme (rather than two independent copies of the same grapheme) on at least some level of the writing architecture, but it is not clear at what stages this comes into play. Existing research on geminate production has largely been in the domain of handwriting, leaving open questions of how they manifest in the more rapid process of typewriting. The present study attempts to isolate the effects of geminate letters on retrieval of spelling from those on articulation.

The interface between spoken and written language has often been described with a dual-route architecture: the pronunciation of written input can be either retrieved from memory via a lexical pathway or constructed from rules of spelling-to-sound conversion via a sub-lexical pathway, and the two pathways operate simultaneously for skilled readers (Coltheart, Curtis, Atkins, & Haller, 1993). Spelling and writing can be explained by simply reversing the direction of the model: the orthography of auditorily-presented words can be retrieved from the lexicon or built from spelling rules (Rapp, Epstein, & Tainturier, 2002). The two routes are thought to converge at the graphemic buffer, a working-memory structure that holds the set of abstract letter identities while they are translated into concrete articulatory patterns (Caramazza & Miceli, 1990).

Most research on the dual-route spelling architecture focuses on handwriting or oral spelling, but more recently it has been applied to typewriting as well (e.g. Pinet, Ziegler, & Alario, 2016). The model's scope largely ends at the graphemic buffer; the process of converting graphemic representations to articulatory plans (for oral spelling, handwriting, or typing) is not specified. Logan and Crump (2011) describe a hierarchical dual-loop model of typewriting, in which an inner loop responsible for translating words to keystrokes is nested within an outer loop that performs the linguistic tasks up to the selection of a word. (Figure 1 illustrates the major elements of this model and of the dual-route model.) The hierarchical model is useful for explaining contrasting effects on typing latencies (intervals between presentation of a word and the initial keystroke) and inter-keystroke intervals, or IKIs, within words. According to

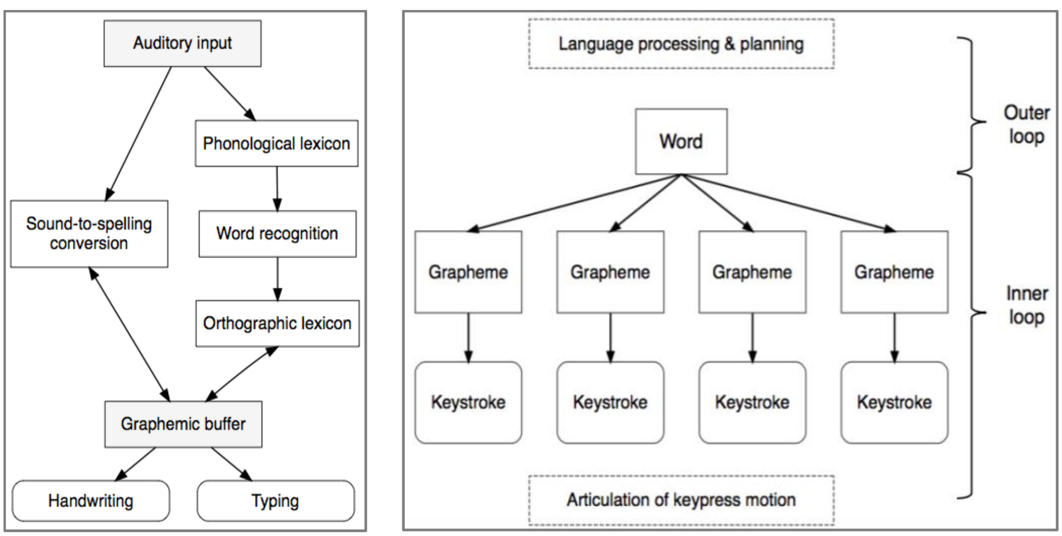

Figure 1. Dual-route model of spelling (e.g. Houghton & Zorzi, 2003; Rapp et al., 2002; as cited in Pinet et al., 2016) and dual-loop hierarchical model of typing (Logan & Crump, 2011); figures adapted from Pinet et al. (2016)

Logan and Crump, latency is a measure of the time taken to perform all outer-loop processes plus the time taken to execute the first keystroke, while inter-keystroke intervals measure the rate of processing in the inner loop.

Pinet et al (2016)'s findings included such contrasting effects: Word frequency, generally considered a lexical feature, affected latencies in a typing-to-dictation task, while bigram frequency, a sub-lexical feature, affected the IKIs. Pinet et al. advocate for an incorporation of elements of both models, with the dual-route architecture explaining most of the outer loop's effects. However, their results also included syllable-level effects: a trend of longer IKIs at syllable boundaries, and an inhibitory effect of syllable sound-spelling consistency[1] on IKIs. The dual-loop model only predicts word-level and grapheme-level effects (observed via latency and IKIs respectively), and includes no processes that would operate over syllable units. To account for this, Pinet et al. argue that there must be some processing level in between the inner and outer loops of a hierarchical model, such as the graphemic buffer.

The nature of geminates within a hierarchical model was investigated by Sternberg, Knoll, and Turock (1990), who concluded that geminate letter pairs are produced in a single unit, or "subprogram," in the inner loop. As defined by Sternberg et al., a subprogram ordinarily consists of the motions and feedback involved in typing a single letter, but their results showed that the two keystrokes of a doubled letter pair are executed by a single subprogram. When participants

[1] Sound-spelling consistency by syllable was calculated as the proportion of "friends" (words with the same syllable, matching both orthography and phonology) out of all French words with the same syllable by orthography.

typed short nonsense letter sequences, IKIs varied by length of the letter sequence and by position within the sequence — but neither factor had an effect on the intervals between the two letters of a doubled pair. This implies that these intervals, unlike all other IKIs, do not correspond to a transition between subprograms.

Further evidence for the representation of geminates as a single doubled grapheme, rather than as two independent copies of the grapheme, is found in the distribution of errors in patients with acquired dysgraphia (impairment in written language processes). Patients' errors on words containing geminates virtually never resulted in the separation of the two letters of the geminate, but very frequently involved geminating the wrong letter, indicating that gemination is a property applied to single graphemes (Caramazza & Miceli, 1990; McCloskey, Badecker, Goodman-Schulman, & Aliminosa, 1994; Tainturier & Caramazza, 1996). For example, in Tainturier & Caramazza (1996), patient errors on the target *ra**bb**it* included productions like *rattib* or *rabitt* far more often than *ra**b**ti**b***. This pattern of geminate pairs being treated as a singular doubled grapheme is not mirrored in other consonant clusters (e.g. *ta**rg**et*), digraphs (e.g. *ro**ck**et*, *bro**th**er*), or non-adjacent duplicate letters (e.g. ***c**a**c**tus*) (Tainturier & Caramazza, 1996). The geminate errors can therefore be best explained as a separation of a doubling feature from the single grapheme to which it applies.

McCloskey et al. (1994) also pointed to the lack of "tripling" in suffixation (e.g. *fu**ll*** + ***l**y* does not produce **fu**lll**y*) as evidence that letter quantity must be specified independently of letter identity: If the geminate ***ll*** in *fu**ll*** were represented as an independent sequence of two copies of the grapheme ⟨l⟩, then the affixation of -*ly* would produce the triplet; the reduction of the triplet to a doublet implies that the doublet was already represented in the quantity tier as a doubling marker linked to a singular ⟨l⟩.

This study aims to examine this theorized quantity tier in more depth and how it operates within models of typed word production. Kandel, Peereman, & Ghimenton (2013, 2014) interpreted their results as evidence of cascading effects of geminates through the dual-route architecture, from central to peripheral processes; the present study attempts to validate this claim by investigating geminate effects on planning and articulation. Results from the present study are compared with predictions from the dual-route hierarchy and dual-loop models in keeping with Pinet et al.'s (2016) recommendations for combining the two.

A challenge of the present study is differentiating geminates' influence on the different stages of production. In particular, it is important to isolate any purely motor effects, because the motion of pressing the same key twice on a keyboard is inherently different from that of pressing two different keys in succession. Previous research has examined the timing of entire utterances to isolate planning and articulation stages: Fink, Oppenheim, and Goldrick (2018) examined both lexical and non-lexical effects on spoken picture-naming, and found that lexical factors' effects on word duration were present even after controlling for effects on latency. This indicated that the lexical effects continued to emerge after response initiation, when retrieval should be finished. Scaltritti et al. (2016) studied typed picture-naming in Italian and also found that lexical effects emerged both before and after response initiation, but that orthographic and sub-lexical effects were only observed as starting after response initiation. The present research attempts to isolate the articulation stage further with a triple-production task: a single trial constitutes typing the target word three times in immediate succession (see e.g. Sausset, Lambert, & Olive, 2013). Any overlap of the retrieval and articulation stages should be restricted to the initial production; by the time of its completion, retrieval must obviously be complete. The following two productions are then a more isolated measurement of the articulation stage – the motor plan and graphemic buffer are still active, but should not be interacting with the earlier stages in the writing architecture.

The present study is a partial replication of research conducted by Kandel et al. (2013, 2014) on the handwritten production of geminates in English and French words, respectively. In both studies, words with geminates occupying the third and fourth letter positions (e.g. *di**ss**ipate*) were matched with controls having the same first three letters (e.g. *di**sg**race*) (2013), excepting one geminate-control pair in the 2013 study and two in the 2014 study whose geminates occupied the fourth and fifth positions. Additionally, all words in the 2014 study had a syllable boundary at the geminate position.

Kandel et al. compared latencies, letter-stroke durations, and inter-letter intervals for the first four letters in handwritten production of the geminate and non-geminate words (2013, 2014). The two studies yielded contrasting results: latencies, stroke durations at each of the first four letters, and durations of each of the first three inter-letter intervals were all decreased by the presence of a geminate in English (2013), but increased by the presence of a geminate in French (excluding the third inter-letter interval, which in French was not significantly affected) (2014). Opposite effects notwithstanding, the authors drew largely the same conclusions from both studies: Geminates are represented with a letter quantity feature which either increases or decreases cognitive load, and this effect cascades from central to peripheral processes, acting on all stages from spelling retrieval through motor execution (Kandel et al., 2013, 2014).

While the contradictory effects could be a true reflection of differences between English and French, limitations in the design call the studies' power into question. The samples of both participants and stimuli were small: the 2013 study included twenty participants and twenty-eight words, and the 2014 study had twenty-seven participants and thirty-two words (Kandel et al.). In addition, the stimuli were imperfectly matched; word pairs seem to have been strictly matched only by the first three letters. The full sets of geminate and control words were matched on word frequency, length in letters, orthographic neighborhood, and bigram frequency at the geminate position (2013, 2014), as well as length in syllables (2013) and age of acquisition (2014). The *p*-values for these characteristics from *t*-tests between the geminate and control sets were not significant (Kandel et al., 2013, 2014), although the 2013 study did not provide any figures or *p*-value for length in syllables despite listing it as a matching characteristic. A *t*-test is not likely to be powerful enough with such small samples, especially if the characteristics are not normally distributed (length in letters is almost uniformly distributed in these sets, and word frequency is strongly skewed).

A stronger set of stimuli would have included strict matching of controls to geminates at least on discrete values like length; both studies included several pairs of mismatched lengths (e.g. *dissipate* and *disgrace*, which are unequal in letter count, syllable count, and morpheme count) (Kandel et al., 2013). Kandel et al. also neglected to match by phonological characteristics, even the phonology of the matched letters: for example, the 2013 study includes the pair (*ma**rr**ow*, *ma**rt**ial*), which share the initial three letters *mar*- but with different phonologies (/mæɹ/ vs. /mɑɹ/); recall that Pinet et al. (2016) reported effects of sound-spelling consistency on IKIs.

The size of the stimuli set was likely influenced by the need to minimize time and fatigue, since handwriting is slower and more effortful than other forms of word production. Handwriting also necessitated a possible threat to validity: participants were required to lift the pen fully off the writing surface between letters in order to produce unambiguous inter-letter intervals. They were allowed a practice period to become accustomed to this motion (Kandel et al., 2013, 2014), but it still calls the results into question if participants were not able to use their natural handwriting style.

The present study uses a typewriting task instead of a handwriting task in order to avoid the limitations of Kandel et al. (2013, 2014). Recording inter-keystroke intervals (IKIs) does not require participants to change their natural typing motions, and the relative quickness of typing allows for far more stimuli to be used with a lesser risk of fatigue effects (N = 230). With a larger number of stimuli and the ability to test more participants ($N = 42$), the present study is able to used a mixed regression model instead of an ANOVA, comparing IKIs on the level of trigrams (sequences of three adjacent letters) rather than paired words. A key benefit of the mixed model approach is the inclusion of control variables as predictors in the regression instead of using them to form matched pairs. Stimuli include words with geminates at varying positions, allowing for examination of geminate effects without confounding by letter position; in the same vein, the entire word is included in the analyses rather than only the letters leading up to the geminate.

As discussed above, the experimental task further departs from a replication of Kandel et al.'s (2013, 2014) by including a triple production of each stimulus. In addition, the stimuli were presented auditorily, following Bonin, Méot, Lagarrigue, and Roux's (2015) findings that variables influencing the sub-lexical pathway of spelling retrieval only affected latencies in a spelling-to-dictation task, not a direct visual copy task.

The present study was designed to investigate the effects of geminate letters on the latencies and IKIs of typed word production. By analyzing latencies and IKIs of triple-productions of words with geminates in a variety of positions, we hope to identify effects in the planning stages to confirm a doubled-grapheme representation on a quantity tier.

Method

Participants

Participants ($N = 42$) were students in one of two undergraduate psychology courses (Introduction to Psychology and Statistics for the Behavioral Sciences) at Tufts University. Participants were compensated with credit towards a research requirement included in the two courses. Their ages ranged from 18 to 21 years (*M* = 18.79, *SD* = 0.83). Sixteen participants were male (38%); twenty-six were female. Nearly all were right-handed (95%). Participants were required to be native or near-native English speakers (exposure to English through at least age 5), have normal or corrected-to-normal hearing, and be able to use a computer mouse and keyboard. Only participants who reported meeting these requirements in an earlier prescreen were able to sign up for the study.

Stimuli

The stimuli consisted of audio recordings of isolated words (*N* = 230) spoken by Mac OS X's text-to-speech synthesizer (using the "Samantha" system voice, a female American-accented voice) compressed to a sample rate of 16 kHz.

Half of the presented words (*N* = 115) contained geminate consonants (e.g. *account*), while the other half (*N* = 115) did not contain geminates (e.g. *achieve*). The full list of stimuli can be found in the Appendix. Words were sourced from the English Lexicon Project (Balota et al., 2007). The selection contained only words with the following characteristics:

- Length = 7 letters
- Monomorphemic
- Raw frequency in HAL corpus ≥ 25
- Mean word-naming accuracy in the ELP's data ≥ 0.92 (first quartile)
- Mean lexical decision accuracy ≥ 0.8

Words were excluded if they contained doubled vowels (e.g. *ch**ee**tah*). Proper nouns were excluded, as were words with homophones or near-homophones (e.g. *glisten* was excluded because of potential confusion with *listen*). These criteria yielded 131 words with geminates and 469 words without geminates. From these, words were also excluded if they were subjectively judged to be unfamiliar, difficult to spell, or otherwise confusing (e.g. *scissor* was excluded because its plural form *scissors* is more common), leaving 120 geminate words and 390 non-geminate words..

The set of non-geminate words were selected from the sample of 390 by randomly generating one hundred possible samples of 120 words and tallying the bigrams in each sample. Each sample's bigram counts were correlated with the bigram counts of the geminate word set (excluding geminate bigrams from the correlation), and the sample with the highest correlation (0.82) was selected.

After generating text-to-speech audio files, another 5 geminate words and 5 non-geminate words were excluded because of incorrect or confusing pronunciation in the text-to-speech, leaving the final samples of 115 geminate words and 115 non-geminate words. All participants were presented with all 230 stimuli. Table 1 shows the characteristics of the geminate and non-geminate stimuli lists.

Apparatus

The experiment was presented on a 21.5-inch iMac desktop computer with a program built with PsychoPy software (version 1.85.6, Peirce et al., 2019). Participants heard the stimuli through standard over-ear headphones and typed

Table 1

Stimuli characteristics provided by the English Lexicon Project (Balota et al.,

	Geminate words		Non-Geminate Words	
	M	*SD*	*M*	*SD*
Raw Lexical Frequency	13658	39125	17205	41508
Log Lexical Frequency	7.66	1.90	8.15	1.79
Mean Naming Accuracy	.990	.020	.992	.019
Mean Lexical Decision Accuracy	.952	.046	.953	.047
Bigram Frequency	1801	632	2011	648

their responses on an Apple A1644 keyboard[2]. The iMac's screen refresh rate was 60Hz, which limits the temporal precision of the results; the PsychoPy program presented stimuli and checked for responses with every new screen frame (every ~16.667 ms).

Procedure

Participants received a brief verbal explanation of the experimental task before responding to demographics questions on the computer. The experiment script began with detailed written instructions presented on the screen. Participants then completed four practice trials of six-letter words not included in the experiment stimuli. During the practice trials, the correct response was displayed on the screen after each trial in the following form: *"You should have typed exactly this: yellow (space) yellow (space) yellow (space)."*

Participants completed 230 trials in a randomized order. The trials were

[2]The A1644 Keyboard is the standard wireless keyboard for recent iMac models. It is "chiclet-style," meaning it has very low, flat keys.

split into six blocks; after 38 or 39 trials were completed a screen was displayed instructing participants to take a break and indicate on the keyboard when they were ready to continue. The experiment's runtime was roughly 40 minutes.

All visual stimuli were centered on the screen, displayed in white on on a 50% gray background and scaled to 10% the height of the screen. A trial began with a fixation cross on the screen, which blinked in unison with a 440-Hz audio tone to signal the coming audio stimulus. Participants responded to the audio stimulus by typing the spoken word three times in succession, pressing the spacebar after each repetition. Their responses appeared in lowercase on the screen as they typed and were cleared when the spacebar was pressed. Any

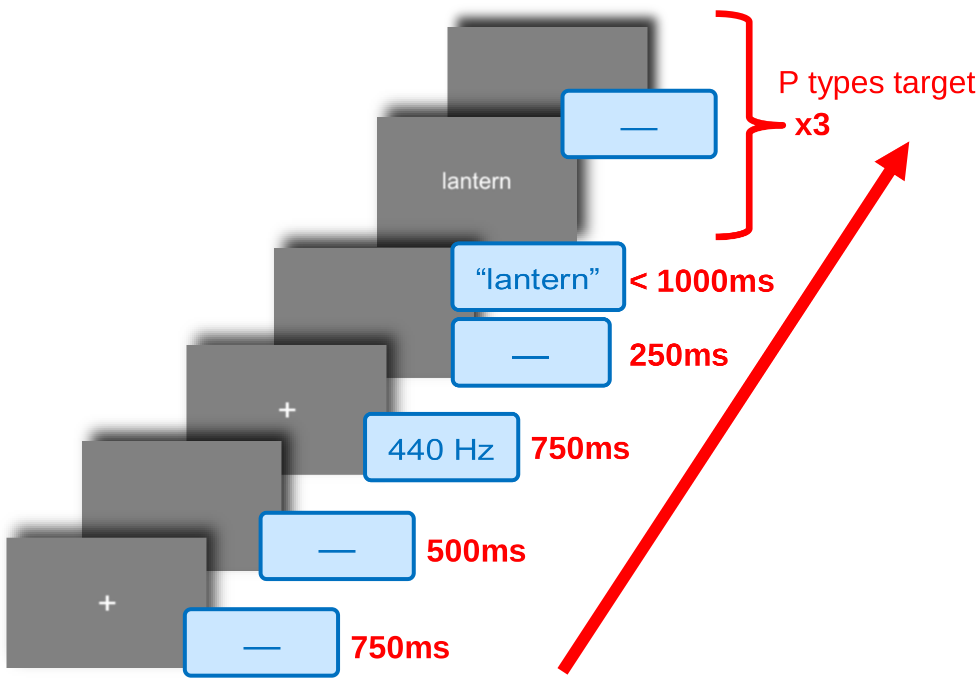

*Figure 2***.** Graphical representation of the screen (grey) and audio (blue) throughout a trial with the word *lantern* as the target, starting in the bottom left and continuing towards the top right.

keypresses other than letters and the spacebar were not displayed or recorded. The trial would not end until the participant pressed the spacebar for the third time, at which point the next trial would immediately begin. (Correct answers were not displayed as they were for the practice trials.) Figure 2 presents a graphical depiction of a trial. After 230 trials were completed, the participant was debriefed and given credit.

Analysis

Typing times were used to build a typing-time corpus following the method proposed by Cohen Priva (2010): inter-keystroke intervals (IKIs) are compared on the basis of three-letter sequences, or trigrams. (For example, the lag between *c* and *h* in the word *achieve* is compared to all other *c-h* lags in the context of the trigram *c-h-i*, which includes *c-h* in *chipper* but is distinct from *c-h* in *chapter*.) Cohen Priva used the median of each trigram's IKI throughout the corpus as a baseline, or expected, value, and the ratio between actual and expected IKIs as the dependent variable for analysis. (Medians were used rather than means because of the non-normal distribution of IKIs.) Trigrams were recommended by Cohen Priva (2010) because of existing research showing that skilled typists usually initiate a keystroke while the preceding keystroke is still in progress, and the articulation of a keypress is therefore influenced by the keys preceding and following it. By evaluating IKIs in the context of the preceding and following keys, one can better control for simple motor constraints relating to the position of keys on a keyboard in the analysis of spelling processes.

The present study's corpus is more limited in scope than that of Cohen

Priva's (2010) English corpus, which includes typing data for a random sampling of thousands of frequent English words; the present study includes only 230 words and therefore contains some trigrams that only appear in the context of one word (e.g. *abdomen* is the only word used containing the trigram *a-b-d*). To avoid confounding expected IKI values by word, the researcher calculated expected values based off of both the present study's corpus and Cohen Priva's English corpus.

Due to the limits on temporal precision imposed by the screen's 60Hz refresh rate, IKIs were linearly transformed to the unit of frames rather than milliseconds (e.g. a recorded IKI of 120ms was multiplied by [60 frames / 1000 ms] resulting in 7.2 frames, rounded to the nearest whole number of frames).

Following Cohen Priva (2010), all lags from incorrectly-typed words and interrupted words (i.e. containing an IKI of more than 500 ms, or 30 frames, after the first key) were excluded. (Only individual productions were excluded in this fashion; a mistake or interruption during one of the three productions of a word did not lead to the exclusion of the other two productions.) Cohen-Priva also excluded all word-initial keystrokes and the top and bottom five percentile of remaining IKIs; the present study separates word-initial keystrokes and word-internal keystrokes into two datasets of latencies (i.e.lag between stimulus presentation and first keypress) and IKIs and excludes only the top and bottom 2.5 percentile of values from each set. Since words were typed three times in succession separated by a spacebar press, the latency dataset only included the word-initial lags of the first production (e.g. *r-SPACEBAR* and *SPACEBAR-c* in "*chipper-SPACE-chipper-SPACE-chipper"* were not included).

Expected values.

Inter-Keystroke Intervals. Expected values for IKIs were calculated as weighted averages of the medians from the present study's corpus (after exclusions) and Cohen Priva's (2010) corpus: for each trigram T,

$$\frac{\left(Mdn(T)_{this\ corpus} \times N(T)_{this\ corpus}\right) + \left(Mdn(T)_{CP\ corpus} \times N(T)_{CP\ corpus}\right)}{N(T)_{this\ corpus} + N(T)_{CP\ corpus}}$$

For example, the trigram *e-c-t* appeared in the present study's corpus 600 times with a median IKI of 10 frames (about 167 ms), and in Cohen Priva's (2010) corpus 243 times with a median IKI of 207 ms, about 12.42 frames. Its expected value is

$$\frac{(10 \times 600) + (12.42 \times 243)}{600 + 243} = \frac{6000 + 3018.06}{843} = 10.70$$

Each production of *e-c-t* in the present study was analyzed in relation to this expected value of 10.70 frames. For word-internal trigrams which did not appear in Cohen Priva's corpus (N=22), the expected value was equal to the median from the present study's corpus. (Each of these 22 trigrams appeared in the present study's corpus at minimum 46 times.) Among word-internal trigrams, the minimum number of appearances in the present corpus was 28. Table 2 summarizes the raw and expected values, as well as the ratios of raw and expected values which constituted the dependent variable in the models described in the "Models" section.

Latencies. As were the IKIs, latencies were evaluated in trigram units. Word-initial trigrams are represented using # to symbolize the beginning of a word, e.g. *#-c-h* for the latency before *chipper*. As this notation suggests, the latencies of course all share a preceding "key" but can be expected to vary by the following key in the same manner as IKIs. Retrieval or "inner-loop" processes are not expected to vary by the second grapheme of a word, but the articulation of the first keypress is still dependent on it and justifies the use of expected values by trigram. The expected latencies of word-initial trigrams were calculated only from the latencies in the present study's corpus. Due to Cohen Priva's (2010) corpus being collected from a visual copy task of individual words and sentences, the latencies from that corpus are not analogous to the latencies in the present study's

Table 2

Summary statistics for IKI values

	Min	Max	Mean	*SD*
IKI (frames)[a]	3	21	8.51	3.49
	(50.0 ms)	*(350.0 ms)*	*(141.89 ms)*	*(58.12 ms)*
1st Production (IKI_1)			8.42	3.48
2nd Production (IKI_2)			8.50	3.46
3rd Production (IKI_3)			8.61	3.52
Expected IKI[b] (frames)	5.19	13.93	8.20	1.74
Ratio[c]	-1.52	1.40	-0.02	0.37
1st Production (IKI_1)			-0.03	0.37
2nd Production (IKI_2)			-0.02	0.37
3rd Production (IKI_3)			-0.01	0.37

[a]Approximate values in milliseconds are shown in parentheses
[b]By-trigram weighted mean of median IKI in present corpus and Cohen Priva's (2010) corpus
[c]Log-ratio of actual value to expected value [log(raw ÷ expected)]

spelling-to-dictation task. (Median latencies from Cohen Priva's corpus correlated with the present study's latencies with $r = .08$; in comparison, median IKIs from Cohen Priva's corpus correlated with the present study's IKIs with $r = .33$.) Without the use of a second corpus, it was necessary to exclude word-initial trigrams which appeared in the present study's corpus very rarely after the previous exclusions. Three trigrams were excluded for this reason, with fewer than 15 appearances: *#-o-f, #-k-n, #-t-w*.

In addition, latencies in the present study were recorded as time since the beginning of the stimulus, and as such were confounded by the acoustic duration of the stimulus. Pinet et al. (2016) accounted for this by entering acoustic duration into the model as a predictor. The current study controls for it more directly by including it in the calculations of expected latency values. Latency values were adjusted by subtracting the stimulus's acoustic duration, such that they measured the elapsed frames from the end of the stimulus presentation rather than the beginning. Medians were calculated on these end-of-stimulus latencies. For example, the acoustic durations of the recordings of *chipper* and *chariot* are 621 ms (37 frames) and 804 ms (48 frames) respectively. A raw latency of 45 frames for any production of *chipper* has an end-of-stimulus latency of $45 - 37 = 8$, while the same raw value for a production of *chariot* has an end-of-stimulus latency of $45 - 48 = -3$. (Negative values indicate that the participant began typing before the end of the stimulus.) The median end-of-stimulus latency for the trigram *#-c-h* (including all values from *chipper, chariot,* and other *ch*-initial words) was 6.

Ratios between the median and actual end-of-stimulus latencies would not be meaningful since the range includes negative numbers; instead, the word stimulus durations were added to the trigram median latencies to produce expected values in the same range as the raw latencies. Continuing the example above, the expected latency value for *chipper* is $6 + 37 = 43$, while for *chariot* the expected latency is $6 + 48 = 56$. Table 3 summarizes the raw, adjusted, and expected values, as well as the ratios of raw and expected values which constituted the dependent variable in the models described below.

Models. Cohen Priva (2010) conducted linear regressions with the actual:expected ratio as the dependent variable. Following Pinet et al. (2016), the present study used an inverse-transformation to approximate a normal distribution of latency ratios (Figure 3) and a log-transformation on the IKI ratios (Figure 4). Also following Pinet et al., separate mixed-effects models were performed on the latencies and IKIs. Within the IKI data, the three productions were analyzed separately with identical model formulas.

Initial models were run with the presence or absence of a geminate in the

Table 3

Summary statistics for latency values

	Min	Max	Mean	*SD*
Latency (raw value)	35	99	52.31	11.97
Adjusted latency[a]	-21	62	7.16	12.51
Expected latency[b]	38.21	68.92	50.16	4.99
Ratio[c]	0.42	1.70	1.00	0.20

[a]Raw latencies adjusted to start from the end of the stimulus duration.
[b]By-trigram median latency + by-word stimulus duration
[c]Inverse ratio of actual value to expected value (expected ÷ raw)

word as the predictor of interest. The IKI models also included trigram position within the word as a predictor. Letter positions of words were coded from 0 (initial letter) to 6; e.g. the letters of *acclaim* were coded as <A>$_0$ <C>$_1$ <C>$_2$ <L>$_3$ <A>$_4$ <I>$_5$ <M>$_6$. Trigrams were assigned the position of their central letters: in *acclaim*, the trigram *c-l-a* (whose IKI value is the lag between the keypresses of *c* and *l*) has position 3, corresponding with <L>$_3$.

Both models included trial number (from 1 to 230) as both a fixed effect and a random intercept, and the ratio of the word's phonological uniqueness point to its length in phonemes (*Carnegie Mellon Pronouncing Dictionary, version 0.07b*, 2015) as a fixed effect. Following both Cohen Priva (2010) and Pinet et al. (2016), by-participant typing speed was included as a fixed effect; random-forest analyses (randomforest package in R, Liaw & Wiener, 2002) were used to select the best participant speed predictors, which were by-participant median latency for the latency model and by-participant median total word duration for the IKI models. Also informed by the random-forest analyses, by-word median latency was included as a fixed effect in the IKI models, but not in the latency model because of the amount of confounding of word-initial trigrams with words. Both models included additional random intercepts of trigram and participant, as well as random slopes by participant; letter identity was included as a random intercept in the IKI models but was excluded from the latency model after an initial test showed it explained no variance. The structure of the latency and IKI models are as follows:

Actual:Expected Latency ~ GeminatePresence + TrialNumber + WordUniquenessPointRatio + ParticipantTypingSpeed$_{latency}$ + (1 + Geminate | Participant) + (1 | TrialNumber) + (1 | Trigram)

Log(Expected:Actual IKI) ~ GeminatePresence + TrigramPosition + TrialNumber + WordUniquenessPointRatio + ParticipantTypingSpeed$_{duration}$ + WordTypingSpeed$_{duration}$ + (1 + Geminate | Participant) + (1 | TrialNumber) + (1 | Trigram) + (1 | Letter)

Due to the possibility of expected values being confounded by geminates, the same models were also run on appropriate subsets of the data: The latency analysis was run on an “intersection” subset, which included only the trigrams which appeared in at least one geminate word and one non-geminate word in the study stimuli (*N* = 30 of 93 word-initial trigrams); and the IKI analyses were run on the subset of IKIs which did not correspond to the letters of a geminate (e.g. in *acclaim*, *a-c-c* and *c-c-l* were excluded from this subset).

Latencies and IKIs from productions of the word *oblique* were excluded from these and the secondary analyses due to its median latency being an outlier (*Mdn* = 55; all other words had a median latency between -9.5 and 35).

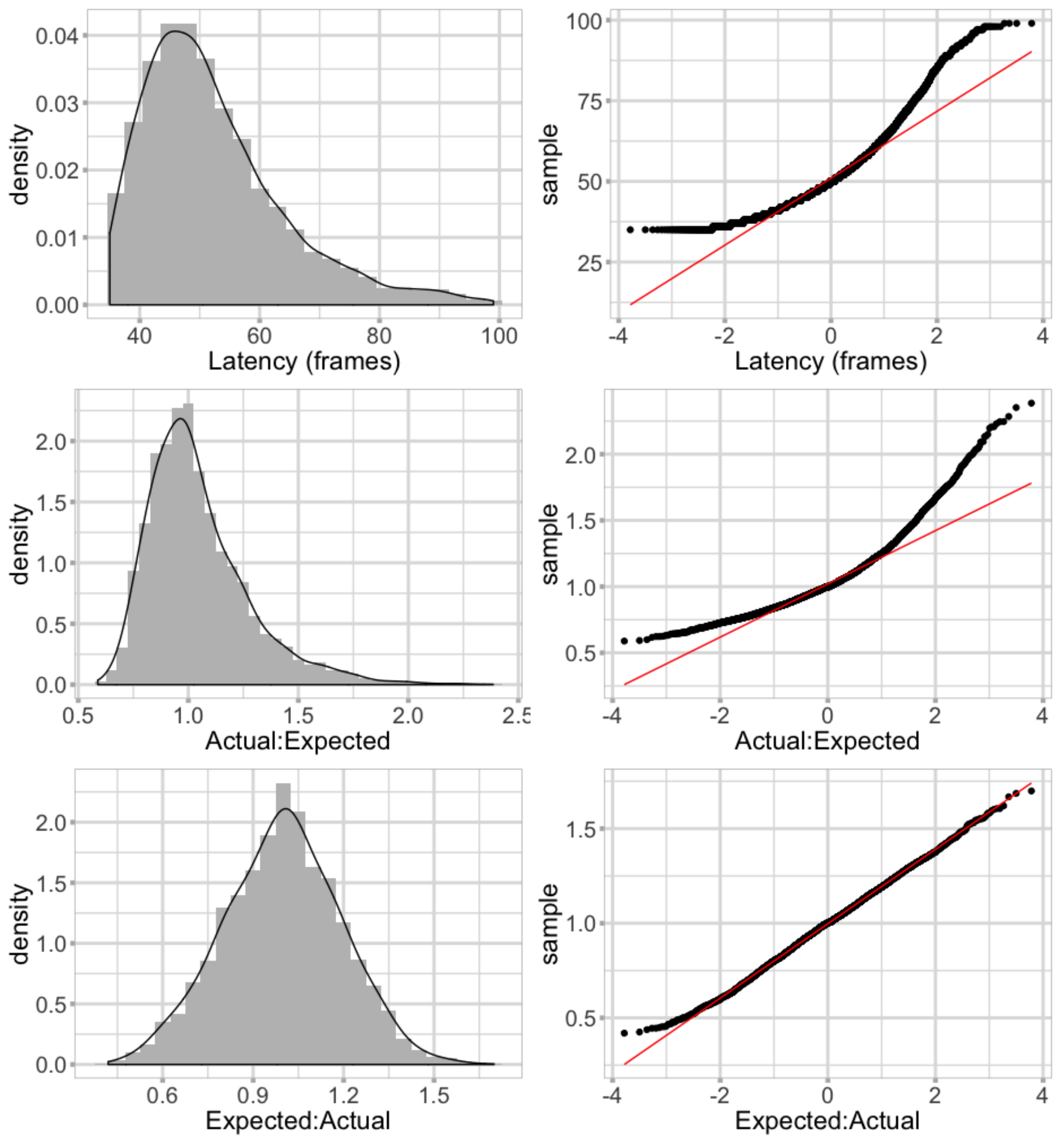

Figure 3. Distribution of response latencies (in rows, respectively: actual values, ratio of actual to expected values, and inverse ratio) as histograms (left) and as quantile-quantile plots comparing the observed distribution on the y-axis to the normal distribution on the x-axis (right).

Additional models were run on the subsets of latencies and IKIs from only the words which contain a geminate. The predictor of interest in these secondary models was the trigram's distance from the geminate in the word. Using the same

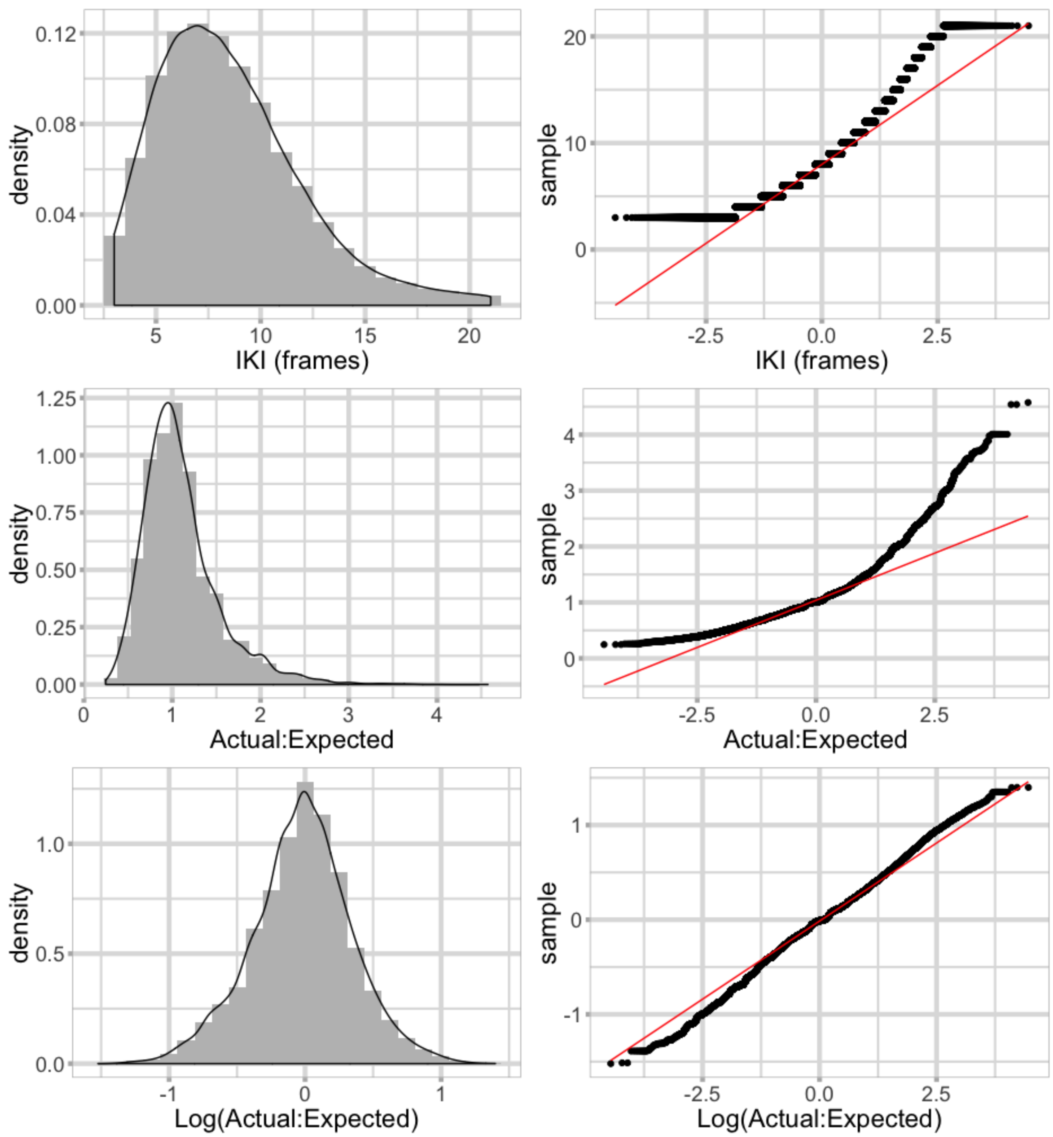

Figure 4. Distribution of inter-keystroke intervals (in rows, respectively: actual values, ratio of actual to expected values, and log-ratio of actual to expected values) as histograms (left) and as quantile-quantile plots comparing the observed distribution on the y-axis to the normal distribution on the x-axis (right).

example of <A>$_0$ <C>$_1$ <C>$_2$ <L>$_3$ <A>$_4$ <I>$_5$ <M>$_6$ from above, the geminate *c*-*c* is located at letters 1 and 2, and the geminate position for the word was therefore coded as 1.5. For each trigram of *acclaim*, distance was then calculated as

TrigramPosition – 1.5. The trigrams *a-c-c* and *c-c-l* in this example have a distance value of -0.5 and 0.5 respectively, and the following trigrams *c-l-a, l-a-i, a-i-m, i-m-#* have distances of 1.5, 2.5, 3.5, and 4.5. In a word like *witness*, where the geminate is at position 5.5, the distance values for non-word-initial trigrams range from -4.5 to 0.5. The full range of distance values in the IKI models is -4.5 to 4.5; for the word-initial trigrams used in the latency analyses, the range is -5.5 to -1.5.

Due to the non-linear, non-monotomic relationship between distance and lags (see Figure 5 and Figure 6), distance was entered into the models as a piecewise term, using natural cubic splines with three degrees of freedom for the latency model and six degrees of freedom for the IKI models (splines package, R Core Team, 2018) (splines package, R Core Team, 2018). The secondary models included all of the same predictors as the initial ones, except for the replacement of the Geminate term with the piecewise distance term and (in the IKI models) the removal of the TrigramPosition term.

Expected:Actual $\text{Latency}_{\text{GeminateWords}}$ ~ DistanceSplines + TrialNumber + WordUniquenessPointRatio + $\text{ParticipantTypingSpeed}_{\text{latency}}$ + (1 + DistanceSplines | Participant) + (1 | TrialNumber) + (1 | Trigram)

Log(Expected:Actual IKI) $_{\text{GeminateWords}}$ ~ DistanceSplines + TrialNumber + WordUniquenessPointRatio + $\text{ParticipantTypingSpeed}_{\text{duration}}$ + $\text{WordTypingSpeed}_{\text{duration}}$ + (1 + DistanceSplines | Participant) + (1 | TrialNumber) + (1 | Trigram) + (1 | Letter)

Latencies and IKIs from productions of the word *address* were excluded from these models because it includes two geminate pairs, which precluded an unambiguous calculation of distance values.

There is as yet no consensus on the most reliable means of significance testing for mixed-effects models. Research by Luke (2017) compares the most

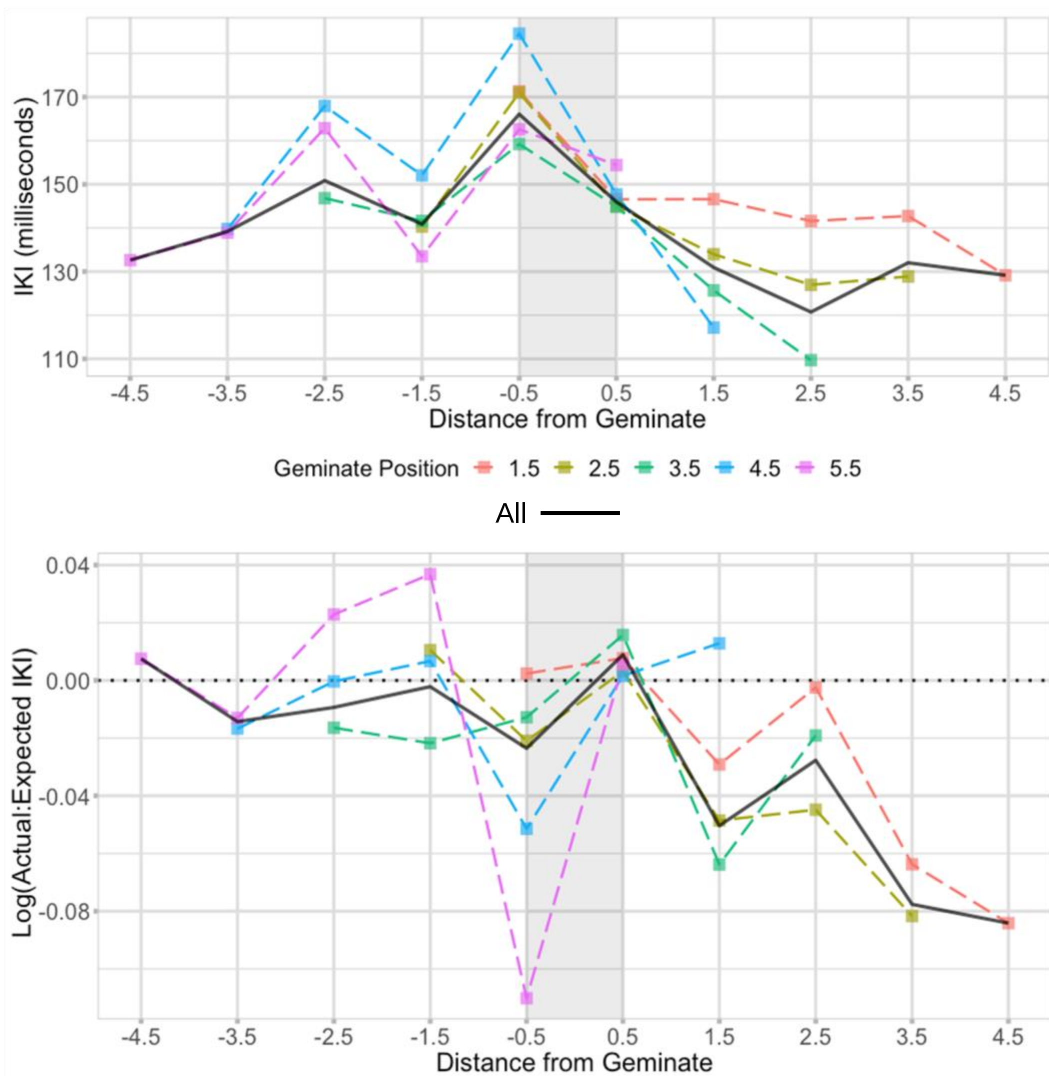

Figure 5. Mean actual IKIs (top) and log-ratios of IKIs (bottom) by position relative to the geminate of the word. Dashed lines show the trends grouped by the geminate's absolute position in the word.

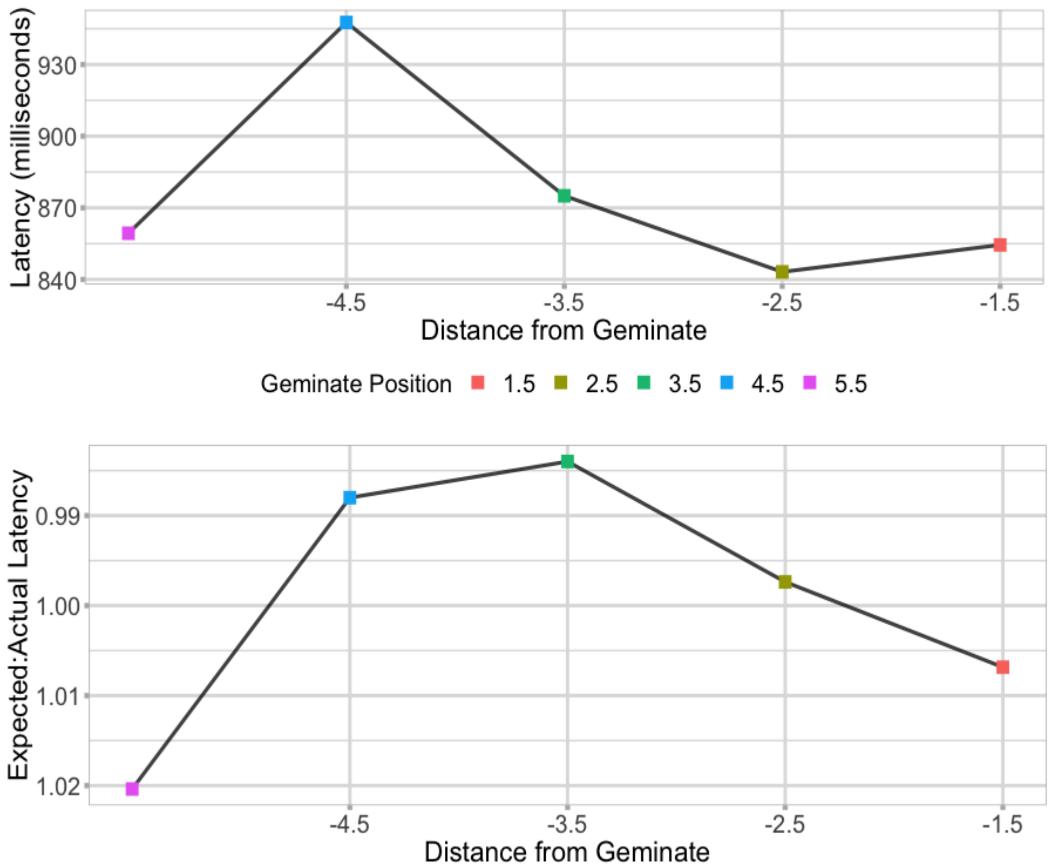

Figure 6. Mean actual latency (top) and inverse-ratios of latency (bottom) by trigram position relative to the geminate of the word. The vertical axis of the ratio graph is reversed to correct for the inverse transformation of the dependent variable: an upward slope indicates an increase in actual latency relative to expected latency.

popular choices: assuming a normal distribution of *t*-values; likelihood-ratio tests; parametric bootstrapping; and computing *p*-values with either of Satterthwaite's or Kenward-Roger's methods for approximating degrees of freedom. The use of *p*-values on a model fitted with reduced maximum likelihood (REML) produced lower Type I error rates (around .05) than either of the first two options, and was less sensitive to sample size than parametric boostrapping. With these results in mind, the present study’s models were fitted with REML and tested for significance with R’s lmerTest package (Kuznetsova, Brockhoff, & Christensen,

2017) using Satterthwaite's method.

Results

Presence of Geminate

Presence or absence of a geminate in the produced word had no significant effect on the inverse latency ratio (*ß* = -.02; *p* = .11), as illustrated in Figure 7 (see Table 4 for the full results of the regression model). The raw latencies were significantly shorter for words with geminates (*M* = 51.8, *SD* = 12.1) than words without geminates (*M* = 52.84, SD = 11.83), *t*(6424.59) = 3.56, *p* < .001, but this trend did not carry through into the ratio of actual to expected values, presumably

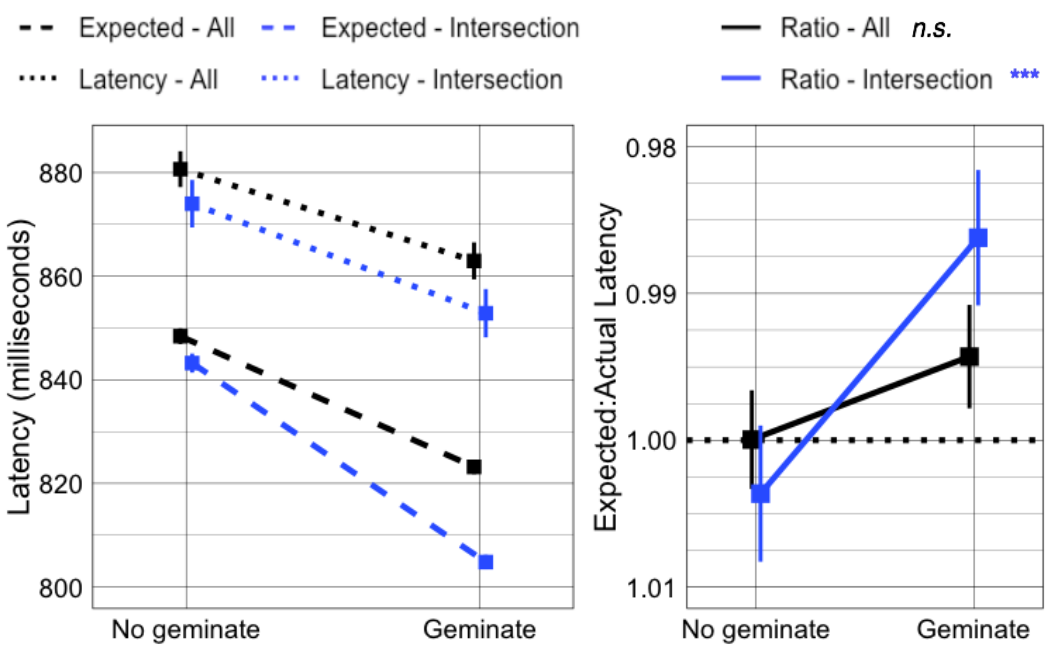

Figure 7. Mean raw latencies and expected values (left) and inverse-ratios (right) for words with and without geminates. The Intersection series includes only data from trigrams that appear in both geminate words and non-geminate words. The vertical axis of the ratio graph is reversed to correct for the inverse transformation of the dependent variable: the upward slope indicates an increase in actual latency relative to expected latency.

because the differences were absorbed into the calculation of expected values. When the model was run on the "intersection" subset (the latencies of trigrams which appeared in both geminate and non-geminate words), the inverse latency ratios were significantly lower for geminate words than for non-geminate words ($ß = -.05$; $p < .001$). (Recall that lower inverse-transformed ratios correspond to *higher* actual latencies relative to expected latencies, so this trend is in the opposite direction of the trend in raw latencies.)

Geminate presence did have significant effects on the log IKI ratios of all three word productions, but the effect reversed direction between the initial production and the last two: As shown in Figure 8, IKIs were longer relative to

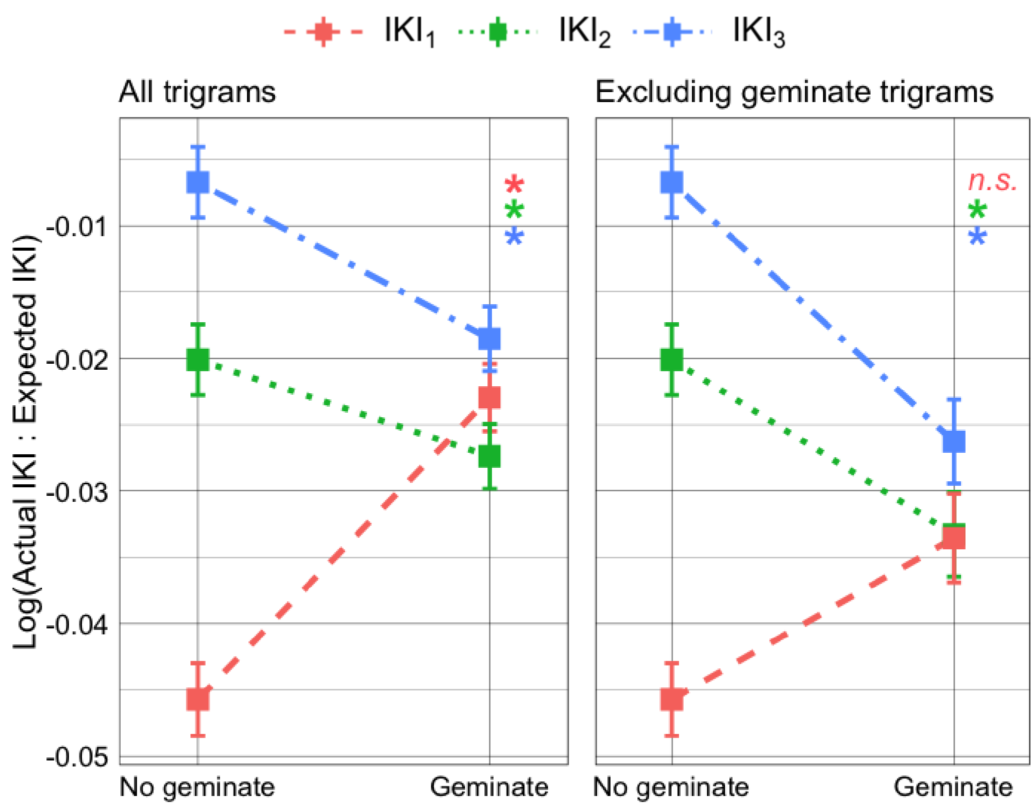

Figure 8. Mean IKI ratios for trigrams within geminate and non-geminate words across the three productions. The right side shows the mean ratios when the geminate trigrams themselves (e.g. *a-c-c* and *c-c-l* of *acclaim*) are excluded.

the expected value when a geminate was present in the initial production (IKI_1), but shorter when a geminate was present in the second (IKI_2) or third (IKI_3) production (see Table 5). However, when the models were re-run excluding the geminate letters, the effect of geminate presence was no longer significant for IKI_1

Within words containing a geminate, the analysis of latencies showed a significant non-linear effect of distance from the geminate based on a Type III ANOVA: $F(3, 93) = 3.23$, $p = .03$ (see Table 6 for the regression coefficients of the cubic splines). Within the IKIs, the effect of distance from the geminate was significant for IKI_1 and IKI_3, but not IKI_2 (IKI_1: $F(6, 55) = 5.26$, $p < .001$; IKI_2: $F(6, 94) = 1.81$, $p = .10$; IKI_3: $F(6, 89) = 3.77$, $p = .002$; see Table 7 for spline coefficients). The marginal effects of distance for each model are illustrated in Figure 9.

Table 4

Results of the initial latency analyses

	B	*ß*	*SE(ß)*	*t*	*p*
All latencies					
(Intercept)	0.984			43.12	< .001 ***
Geminate in word	-0.008	-0.02	0.013	-1.64	.11
Trial number	0.0004	0.15	0.012	11.75	< .001 ***
Typing speed[a]	-0.018	-0.48	0.016	-30.03	< .001 ***
Word uniquness point ratio[b]	0.051	0.03	0.011	2.62	.009 **
Intersection[c]					
(Intercept)	0.941			27.11	< .001 ***
Geminate in word	-0.020	-0.05	0.015	-3.36	< .001 ***
Trial number	0.0005	0.16	0.015	10.73	< .001 ***
Typing speed[a]	-0.018	-0.48	0.017	-27.56	< .001 ***
Word uniquness point ratio[b]	0.086	0.04	0.015	2.82	< .001 ***

Note. *B* = nonstandardized coefficient estimates, *ß* = standardized estimates, *SE* = standard error, *p*-values based on degrees of freedom calculated using Satterthwaite's method

[a]Median response latency by participant (between all words)
[b]Phonological uniqueness point by phoneme number, divided by total number of phonemes (*CMU Pronouncing Dictionary*, 2015)
[c]The Intersection subset includes only latencies of trigrams that appear in both geminate words and non-geminate words within the stimuli

*$p < .05$, **$p < .01$, ***$p < .001$

Table 5

Results of the initial IKI analyses

	Initial Production (IKI_1)					Second Production (IKI_2)					Third Production (IKI_3)				
	B	*ß*	*SE(ß)*	*t*	*p*	*B*	*ß*	*SE(ß)*	*t*	*p*	*B*	*ß*	*SE(ß)*	*t*	*p*
All trigrams															
(Intercept)	-0.891			-17.44	< .001***	-0.91			-21.73	< .001***	-0.841			-18.47	< .001***
Geminate in word	0.014	0.02	0.007	2.61	.01 *	-0.012	-0.02	0.008	-2.19	.03 *	-0.017	-0.02	0.009	-2.62	.01 *
Trigram position	-0.013	-0.06	0.007	-8.53	< .001***	-0.001	-0.003	0.007	-0.44	.66	-0.006	-0.03	0.008	-3.52	< .001***
Trial number	-0.0002	-0.03	0.005	-5.95	< .001***	-0.0003	-0.06	0.005	-13.01	< .001***	-0.0003	-0.06	0.005	-11.94	< .001***
Typing speed[a]	0.017	0.29	0.014	20.24	< .001***	0.017	0.29	0.011	27.21	< .001***	0.017	0.29	0.012	24.44	< .001***
Word UP ratio[b]	-0.037	-0.01	0.006	-1.99	.047 *	-0.01	-0.003	0.006	-0.57	.57	-0.026	-0.01	0.006	-1.40	.16
Word typing latency[c]	0.002	0.03	0.006	6.13	< .001***	0.002	0.03	0.005	5.66	< .001***	0.001	0.02	0.006	4.23	< .001***
Excluding geminate															
(Intercept)	-0.946			-17.18	< .001***	-0.956			-19.73	< .001***	-0.855			-15.31	< .001***
Geminate in word	0.008	0.01	0.008	1.32	.19	-0.017	-0.02	0.009	-2.36	.02 *	-0.02	-0.03	0.009	-2.76	.01 *
Trigram position	-0.012	-0.06	0.008	-7.07	< .001***	0.0004	0.002	0.008	0.22	.82	-0.005	-0.03	0.008	-3.08	< .001***
Trial number	-0.0002	-0.03	0.006	-5.68	< .001***	-0.0003	-0.06	0.005	-11.13	< .001***	-0.0003	-0.06	0.005	-11.20	< .001***
Typing speed[a]	0.017	0.29	0.015	19.77	< .001***	0.017	0.29	0.012	23.50	< .001***	0.017	0.29	0.015	19.10	< .001***
Word UP ratio[b]	-0.030	-0.01	0.006	-1.44	.15	0.003	0.001	0.006	0.15	.88	-0.024	-0.01	0.006	-1.17	.24
Word typing latency[c]	0.003	0.04	0.006	6.14	< .001***	0.002	0.03	0.006	5.69	< .001***	0.002	0.02	0.006	3.84	< .001***

Note. *B* = nonstandardized coefficient estimates, *ß* = standardized estimates, *SE* = standard error, *p*-values based on degrees of freedom calculated using Satterthwaite's method.

[a]Median response latency by participant (between all words)
[b]Phonological uniqueness point by phoneme number, divided by total number of phonemes; sourced from *CMU Pronouncing Dictionary* (2015)
[c]Median latency by word (between all participants)

$^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$

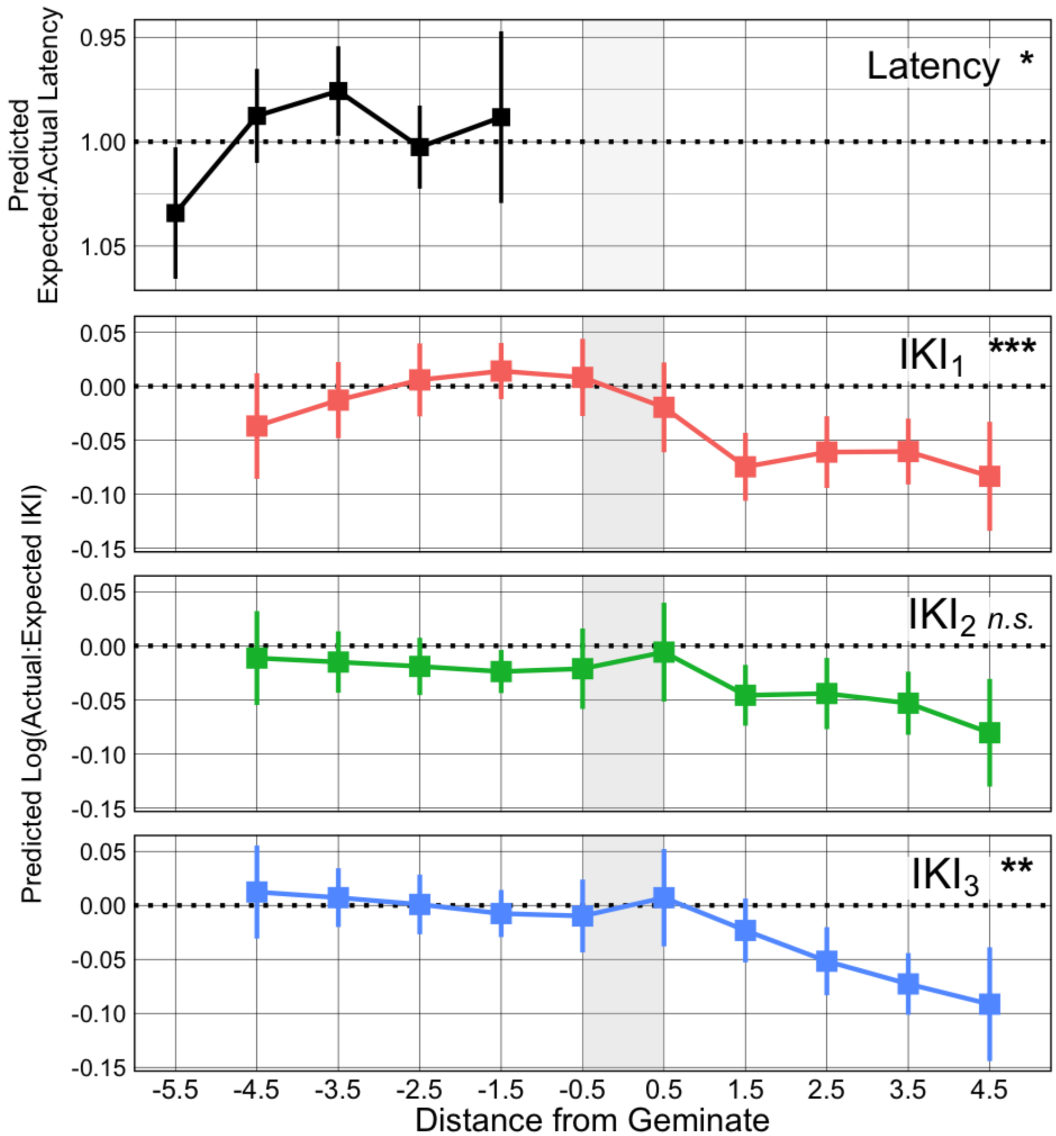

Figure 9. Mean predicted values of the dependent variables in the geminate-word-only models, plotted by distance from the geminate with all other predictors held constant. Confidence intervals are shown as 1.96 × *SE*(predicted). Predicted values were calculated with ggeffects package in R (Lüdecke, 2018). The vertical axis of the latency graph (row 1) is reversed to correct for the inverse transformation of the dependent variable.

Controls

Trigram position showed a general speeding-up of IKIs over the course of a production, but, like distance from the geminate, the effect was only present in

IKI_1 and IKI_3. The relationship is rather non-linear in the second two productions especially, as seen in Figure 10.

Figure 11 illustrates the relationship between by-word median latency and IKI values, which was significant for all three productions. The effects of trial number, participant typing speed, and phonological uniqueness point in the latency and IKI models are shown in Figure 12; trial number showed a significant facilitatory effect across all of the models, indicating improvement with practice, and participant typing speed was by far the strongest predictor in each model ($ß < -.4$ in the latency models; $ß \geq .29$ in each of the IKI models; see Tables 1-4). Phonological uniqueness points closer to the end of the word corresponded to shorter lags relative to expected values in the initial latency analysis and in IKI_1; this effect was not significant in the secondary analyis of latencies (i.e., the analysis of latencies in only the geminate words, using the non-linear distance term), but did remain significant in the secondary analysis of IKI_1. The correlation of by-participant estimated effect sizes (i.e., random slopes by geminate position) and typing speeds was not significant in any of the initial models (Latency: $r = .25$; $p = .11$; IKI_1: r = -.21, p = . 0.19; IKI_2: r = .01, $p = .95$; IKI_3: r = .06, $p = .73$).

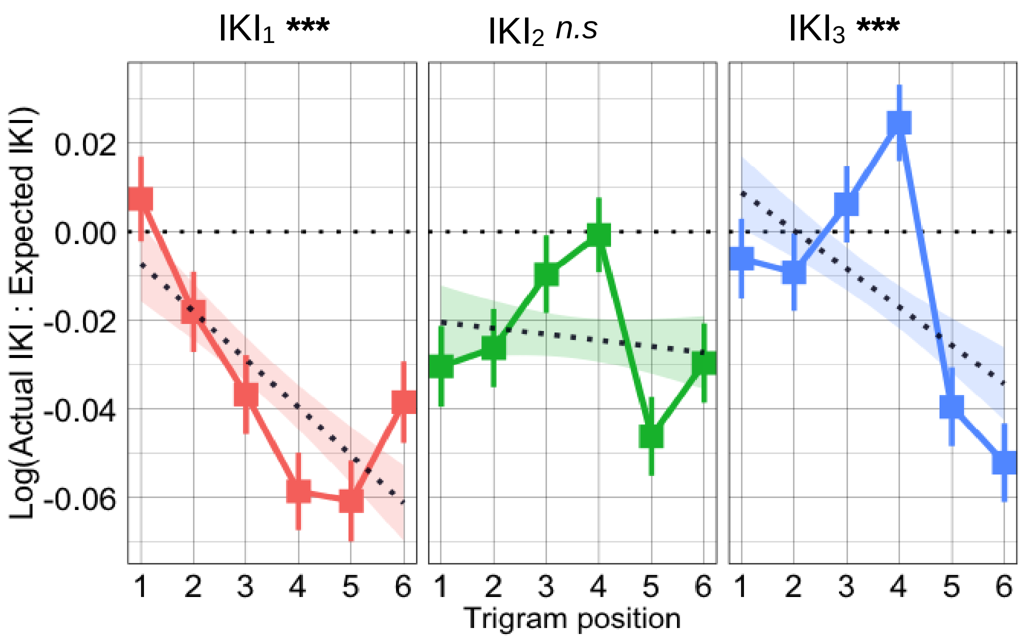

Figure 10. Mean IKI ratio (top) and smoothed conditional means (bottom) by trigram position within the word, grouped by the three productions.

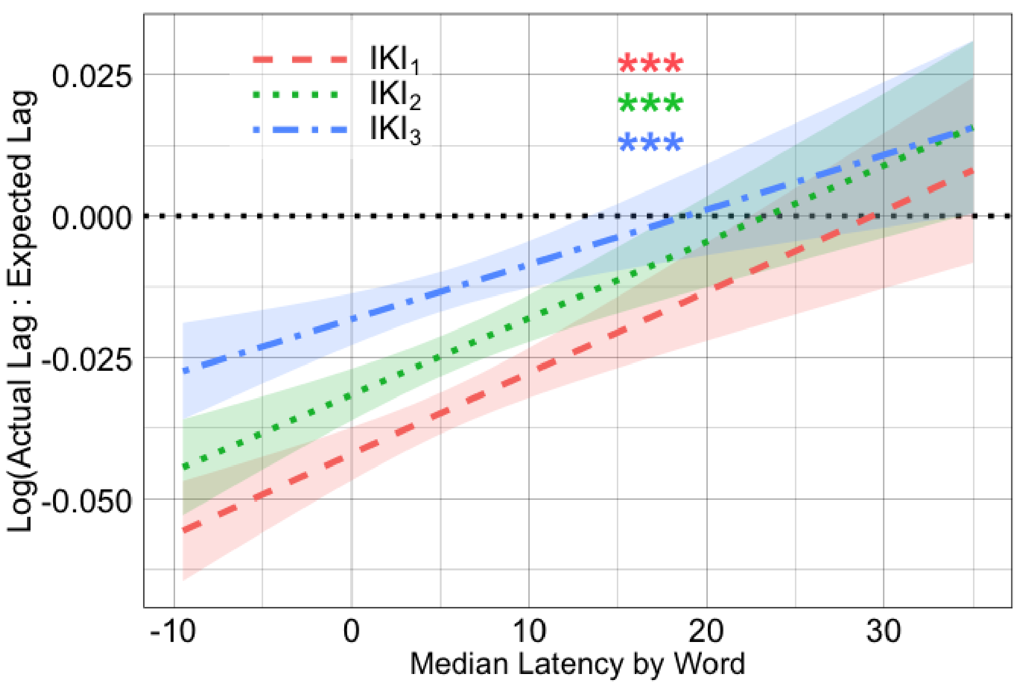

Figure 11. Smoothed conditional means of IKI ratio plotted against median end-of-word latency across all participants, grouped by the three productions.

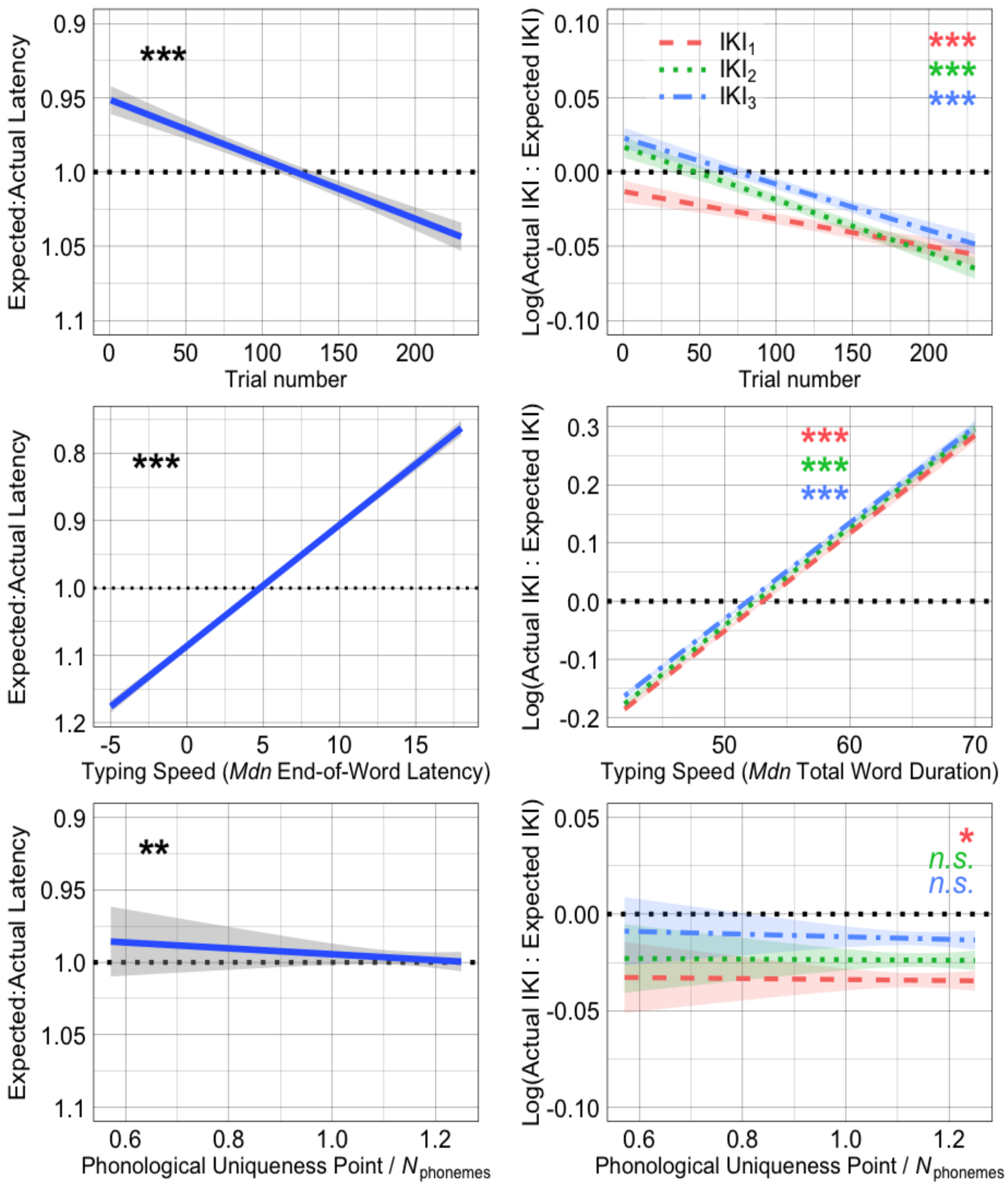

Figure 12. Smoothed conditional means of latency (left) and IKIs (right) by trial number, typing speed, and phonological uniqueness point (*CMU Pronouncing Dictionary*, 2015). The vertical axes of the latency graphs are reversed to correct for the inverse transformation of the dependent variable: the downward slopes indicate a decrease in latency relative to expected latency.

Table 6

Results of the secondary latency analysis including only words with geminates

	B	*ß*	*SE(ß)*	*t*	*p*
(Intercept)	1.006			19.99	< .001 ***
Distance from Geminate[a]					.02[b] *
Spline 1	0.006	0.01	0.026	0.34	0.74
Spline 2	-0.117	-0.08	0.028	-2.92	< .001 ***
Spline 3	0.004	0.01	0.036	0.17	0.87
Trial number	0.0005	0.16	0.017	9.61	< .001 ***
Typing speed[b]	-0.017	-0.45	0.02	-22.76	< .001 ***
Word uniquness point ratio[c]	0.052	0.02	0.019	1.25	0.21

Note. *B* = nonstandardized coefficient estimates, *ß* = standardized estimates, *SE* = standard error, *p*-values based on degrees of freedom calculated using Satterthwaite's method.

[a]Internal splines were set at -3.5 and -2.5; spline 3 is a boundary knot at -1.5.
[b]The *p*-value for the overall predictive strength of distance from the geminate was calculated with a Type III ANOVA test ($F(3, 94) = 3.27$).
[c]Median response latency by participant (between all words)
[d] Phonological uniqueness point by phoneme number, divided by total number of phonemes; sourced from *CMU Pronouncing Dictionary* (2015)

$^*p < .05$, $^{**}p < .01$, $^{***}p < .001$

Table 7

Results of the secondary IKI analysis including only words with geminates

	Initial Production (IKI_1)					Second Production (IKI_2)					Third Production (IKI_3)				
	B	*ß*	*SE(ß)*	*t*	*p*	*B*	*ß*	*SE(ß)*	*t*	*p*	*B*	*ß*	*SE(ß)*	*t*	*p*
(Intercept)	-0.814			-11.98	< .001***	-0.909			-16.31	< .001***	-0.881			-15.27	< .001***
Distance from Geminate[a]					< .001[b] ***					.11[b]					.003[b] **
Spline 1	0.055	0.04	0.028	1.46	.15	-0.011	-0.01	0.03	-0.28	.78	-0.031	-0.02	0.033	-0.70	.48
Spline 2	0.039	0.03	0.044	0.60	.55	0.031	0.02	0.036	0.58	.56	0.008	0.01	0.035	0.16	.87
Spline 3	-0.053	-0.04	0.025	-1.46	.15	-0.046	-0.03	0.023	-1.42	.16	-0.04	-0.03	0.026	-1.07	.29
Spline 4	-0.017	-0.01	0.022	-0.50	.62	-0.012	-0.01	0.023	-0.33	.74	-0.07	-0.05	0.028	-1.70	.10
Spline 5	0.042	0.02	0.035	0.56	.58	-0.046	-0.02	0.028	-0.77	.44	-0.104	-0.05	0.027	-1.81	.07
Spline 6	-0.074	-0.04	0.015	-2.35	.02 *	-0.061	-0.03	0.014	-2.18	.03 *	-0.097	-0.05	0.014	-3.29	< .001***
Trial number	-0.0002	-0.04	0.007	-5.55	< .001***	-0.0004	-0.07	0.006	-10.40	< .001***	-0.0004	-0.07	0.007	-10.80	< .001***
Typing speed[a]	0.016	0.29	0.016	18.41	< .001***	0.017	0.31	0.011	28.23	< .001***	0.017	0.32	0.012	27.25	< .001***
Word UP ratio[b]	-0.092	-0.02	0.008	-3.00	< .001***	-0.015	-0.004	0.008	-0.52	.60	-0.012	-0.003	0.008	-0.39	.69
Word typing latency[c]	0.002	0.04	0.008	4.40	< .001***	0.002	0.03	0.008	4.32	< .001***	0.002	0.03	0.008	3.49	< .001***

Note. *B* = nonstandardized coefficient estimates, *ß* = standardized estimates, *SE* = standard error, *p*-values based on degrees of freedom calculated using Satterthwaite's method.

[a]Internal splines were set at -1.5, -0.5, 0.5, 1.5, and 2.5; spline 6 is a boundary knot at 4.5.
[b]The *p*-value for the overall predictive strength of distance from the geminate was calculated from a Type III ANOVA test (IKI_1: $F(6, 67) = 6.34$; IKI_2: $F(6, 89) = 1.77$; IKI_3: $F(6, 70) = 3.67$)
[b]Median response latency by participant (between all words)
[c] Phonological uniqueness point by phoneme number, divided by total number of phonemes; sourced from *CMU Pronouncing Dictionary* (2015)
[d]Median latency by word (between all participants)

$^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$

Discussion

The present study investigated the production of words with geminate consonants to evaluate their representation at different levels of the written-language architecture. In particular, the study is an attempt to replicate the results from one or both of Kandel et al.'s (2013, 2014) experiments, which demonstrated cascading facilitatory and inhibitory effects of geminates on written word production in English and French respectively. Data was collected from three successive typed productions of geminate words and controls in order to separate out effects on the retrieval stage from effects on articulation. The latencies and inter-keystroke intervals were used to build a typing-time corpus in the style of Cohen Priva (2010), which allowed for the comparison of each trigram to an average from the corpus. Geminates had an inhibitory effect on word latencies, among the word-initial bigrams that were present in both geminate and non-geminate words in the stimuli. This effect varied non-linearly by the geminate's position in the word, seeming most pronounced for geminates near the center of the word.

In contrast, the inter-keystroke intervals (IKIs) of the second and third productions showed a facilitatory effect of geminates — this in turn was contrary to the trend in IKIs of the initial word production. The initial production had slower IKIs in geminate words overall, matching the latencies, but when the IKIs of the geminate letters themselves were removed from the analyses, the rest of the IKIs in the initial production were not sigificantly different depending on geminate. The relationship of IKIs to distance from the geminate within the word was inconsistent: a visual analysis suggests that IKIs get shorter relative to their expected values immediately after the geminate for all three productions, but the non-linear term was not significant for the second production. This effect may have also been confounded by a general trend of shorter IKIs later in the word, which was also signficant for the initial and final productions but not the second.

These results are contrary to the effects reported in Kandel et al. (2013, 2014). In both of those studies, the effects on latencies, letter durations, and inter-letter intervals were in the same direction, which was interpreted as a decreased (2013) or increased (2014) cognitive load cascading from central processes to peripheral processes. One possible explanation for this is that cascading effects do exist in handwriting, but not in typewriting. Being slower, handwriting requires the graphemic buffer to hold information longer, which could result in an overall higher cognitive load during the articulation stage of handwriting that is not present in the articulation stage of typing. This would explain why the letter durations and intervals were susceptible to the same cognitive load effects as the latencies were.

The disparity between latency and IKI effects in the present study prompts a shift in perspective to a hierarchical model. In nonsense strings of letters, Sternberg et al. (1990) found evidence that doubled letter pairs are operated on as single subprograms rather than as copies of the same subprogram. As discussed earlier, the IKIs between geminate letters did not vary by position or sequence length as all other IKIs did (Sternberg et al.). The variance of IKI-ratios between the two letters of the geminate in the present study is in fact less than that of other trigrams; Figure 5 shows that the means are closely clustered around zero, suggesting less variance. This is further supported by examining the individual trigrams' distributions of IKI ratios: the trigrams corresponding to the interval between the two letters of a geminate (*N* = 52) showed less variance than the majority of other trigrams. Figure 13 displays a summary of the standard deviations of ratios for each trigram. This pattern is in line with Sternberg et al's (1990) theory: the intervals between geminate letters are less susceptible to variance arising from any of the factors that affect IKIs between two different letters, because the geminate unit is processed as a standalone "subprogram."

Sternberg et al. (1990) also determined that the total typing duration of a

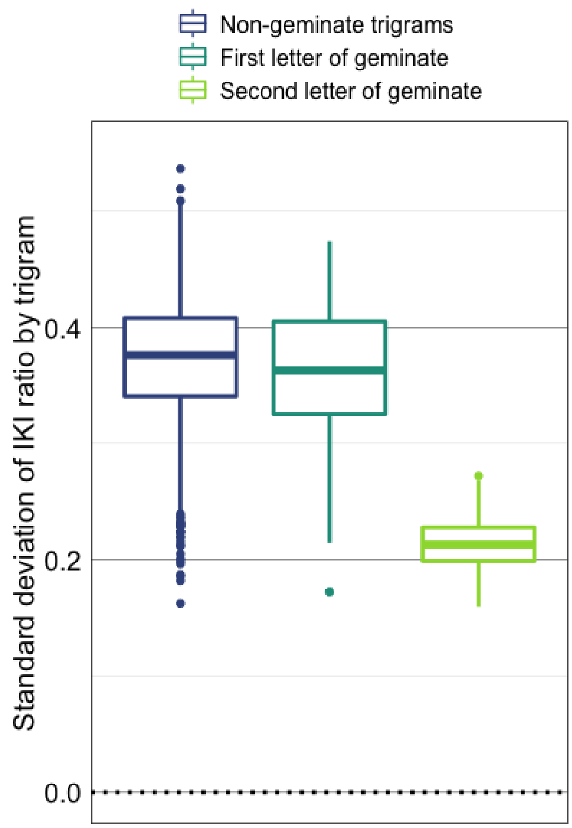

Figure 13. Summary of the standard deviations of IKI ratios for each unique trigram in the stimuli words, grouped by whether the trigram forms part of a geminate.

letter sequence is a function of the number of subprograms rather than the number of letters; a sequence with a doubled letter (e.g. *zxttc*) takes only slightly more time to type than the same sequence with a single copy of the letter (*zxtc*) because it has the same number of subprograms. While the summed duration did not change, the distribution of individual IKIs within the sequence was affected by the presence of a geminate, which Sternberg et al. suggest may be explained by a "temporal modulation" between levels of the hierarchical typing process. Pinet et al. (2016) suggest a graphemic buffer as this modulating component. The facilitatory effect of geminate on the second and last productions in the present study is consistent with this: the geminate words have one fewer subprogram than the non-geminate words, so their total duration and therefore individual IKIs should be shorter.

The lack of facilitation on the initial production may be due to the fact that the present study used real words rather than non-word sequences. The outer loop of copying a non-word sequence will of course not be able to utilize the lexical or sub-lexical routes, but must involve a letter-by-letter identification. Since geminates have an inhibitory effect on latencies, they must add to the cognitive load of some element of the outer loop. The IKIs of the initial production are affected by many of the same factors that act on latency (Fink et al., 2018; Scaltritti et al., 2016), while the following productions' IKIs should not be; it may be that the facilitatory effect of having one fewer subprogram is present for all three productions, but canceled out by the inhibitory effects on the outer loop during the first production.

The boundary conditions of the inhibitory and facilitatory effects should by no means be interpreted as simply the outer and inner loops; for one thing, some components of the outer loop must still be active during the second and third productions. This highlights the potential value in further exploration of a repeated-production task, so that we might determine more specifically which components are and are not active during successive repetitions. Similarly, Sternberg et al.'s (1990) model of subprograms is only defined for articulatory processes in the inner loop, but further research on real-word production could determine whether it extends into a graphemic-buffer-level quantity tier or further up into the dual-route structure.

The implications of these results are unclear, in particular the difference between these and the results of Kandel et al. (2013, 2014). The present study's design made several improvements to Kandel et al.'s, but the added complexity comes at the expense of interpretability. Cascading effects like those reported by Kandel et al. are not evident, but due to the ambiguity of the first production, it is possible that the cascading effects do exist and the statistical analysis was not powerful enough to detect them — and of course it is also possible that they only exist in handwriting. Future research on geminate effects might do well to investigate interactions with modality or even simply with time. In a similar vein, the investigation of (especially typed) word production may benefit from the use of multiple-production designs to better isolate the stages of retrieval and production.

References

Balota, D. A., Yap, M. J., Cortese, M. J., Hutchison, K. A., Kessler, B., Loftis, B., … Treiman, R. (2007). The English Lexicon Project. *Behavior Research Methods*, *39*(3), 445–459. https://doi.org/10.3758/BF03193014

Bonin, P., Méot, A., Lagarrigue, A., & Roux, S. (2015). Written object naming, spelling to dictation, and immediate copying: Different tasks, different pathways? *Quarterly Journal of Experimental Psychology*, *68*(7), 1268–1294. https://doi.org/10.1080/17470218.2014.978877

Caramazza, A., & Miceli, G. (1990). The structure of graphemic representations. *Cognition*, *37*(3), 243–297. https://doi.org/10.1016/0010-0277(90)90047-N

Carnegie Mellon Pronouncing Dictionary, version 0.07b. (2015). Carnegie Mellon University. Retrieved from http://www.speech.cs.cmu.edu/cgi-bin/cmudict

Cohen Priva, U. (2010). *Constructing Typing-Time Corpora: A New Way to Answer Old Questions*. 6.

Coltheart, M., Curtis, B., Atkins, P., & Haller, M. (1993). Models of reading aloud: Dual-route and parallel-distributed-processing approaches. *Psychological Review*, *100*(4), 589–608. http://dx.doi.org.ezproxy.library.tufts.edu/10.1037/0033-295X.100.4.589

Fink, A., Oppenheim, G. M., & Goldrick, M. (2018). Interactions between lexical access and articulation. *Language, Cognition and Neuroscience, 33*(1), 12–24. https://doi.org/10.1080/23273798.2017.1348529

Kandel, S., Peereman, R., & Ghimenton, A. (2013). Further evidence for the interaction of central and peripheral processes: the impact of double letters in writing English words. *Frontiers in Psychology*, *4*. https://doi.org/10.3389/fpsyg.2013.00729

Kandel, S., Peereman, R., & Ghimenton, A. (2014). How do we code the letters of a word when we have to write it? Investigating double letter representation in French. *Acta Psychologica*, *148*, 56–62. https://doi.org/10.1016/j.actpsy.2014.01.002

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software*, *82*(1), 1–26. https://doi.org/10.18637/jss.v082.i13

Liaw, A., & Wiener, M. (2002). Classification and Regression by randomForest. *R News*, *2*(3), 18–22.

Logan, G. D., & Crump, M. J. C. (2011). Chapter one - Hierarchical Control of Cognitive Processes: The Case for Skilled Typewriting. In B. H. Ross (Ed.), *Psychology of Learning and Motivation* (pp. 1–27). https://doi.org/10.1016/B978-0-12-385527-5.00001-2

Luke, S. G. (2017). Evaluating significance in linear mixed-effects models in R. *Behavior Research Methods*, *49*(4), 1494–1502. https://doi.org/10.3758/s13428-016-0809-y

McCloskey, M., Badecker, W., Goodman-Schulman, R. A., & Aliminosa, D. (1994). The structure of graphemic representations in spelling: Evidence from a case of acquired dysgraphia. *Cognitive Neuropsychology*, *11*(3), 341–392. https://doi.org/10.1080/02643299408251979

Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., … Lindeløv, J. K. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods*, *51*(1), 195–203. https://doi.org/10.3758/s13428-018-01193-y

Pinet, S., Ziegler, J. C., & Alario, F.-X. (2016). Typing is writing: Linguistic properties modulate typing execution. *Psychonomic Bulletin & Review*, *23*(6), 1898–1906. https://doi.org/10.3758/s13423-016-1044-3

R Core Team. (2018). *R: A language and environment for statistical computing*. Retrieved from https://www.R-project.org/.

Rapp, B., Epstein, C., & Tainturier, M.-J. (2002). The integration of information across lexical and sublexical processes in spelling. *Cognitive Neuropsychology*, *19*(1), 1–29. https://doi.org/10.1080/0264329014300060

Sausset, S., Lambert, E., & Olive, T. (2013). Flexibility of orthographic and graphomotor coordination during a handwritten copy task: effect of time pressure. *Frontiers in Psychology*, *4*. https://doi.org/10.3389/fpsyg.2013.00866

Scaltritti, M., Arfé, B., Torrance, M., & Peressotti, F. (2016). Typing pictures: Linguistic processing cascades into finger movements. *Cognition*, *156*, 16–29. https://doi.org/10.1016/j.cognition.2016.07.006

Sternberg, S., Knoll, R. L., & Turock, D. L. (1990). Hierarchical control in the execution of action sequences: Tests of two invariance properties. In M.

Jeannerod (Ed.), *Attention and performance XIII: Motor representation and control* (pp. 3–55). Retrieved from https://pdfs.semanticscholar.org/398c/ecf341b9288bfa7c19556c05a7b71371eefc.pdf

Tainturier, M.-J., & Caramazza, A. (1996). The Status of Double Letters in Graphemic Representations. *Journal of Memory and Language*, *35*(1), 53–73. https://doi.org/10.1006/jmla.1996.0003

Appendix

Table A1

Characteristics of geminate stimuli words

Word	Freq[i]	Log Freq[ii]	Acc. LD[iii]	Acc. Nmg[iv]	BG Freq[v]	UP[vi]	*N* Phon.[vii]	UP / *N* Phon.	Duration[viii]
acclaim	889	6.79	0.94	1.00	1015	6	5	1.200	0.804
account	60958	11.02	1.00	0.97	1785	6	5	1.200	0.720
address	221551	12.31	1.00	1.00	2043	4	5	0.800	0.773
apparel	578	6.36	0.97	1.00	1953	4	6	0.667	0.779
applaud	1303	7.17	0.94	1.00	833	6	5	1.200	0.790
appoint	1079	6.98	0.88	1.00	2120	6	5	1.200	0.734
approve	4965	8.51	0.97	1.00	1218	6	5	1.200	0.768
arrange	6050	8.71	0.97	1.00	2544	5	4	1.250	0.814
assault	10485	9.26	0.94	1.00	939	6	5	1.200	0.805
attempt	41637	10.64	1.00	1.00	1917	7	6	1.167	0.778
attract	5510	8.61	0.97	1.00	2042	7	6	1.167	0.770
barrier	7209	8.88	0.97	0.96	2435	6	5	1.200	0.705
battery	27473	10.22	1.00	1.00	2804	6	5	1.200	0.656
billion	20698	9.94	0.88	1.00	2255	7	6	1.167	0.697
bizarre	5944	8.69	1.00	1.00	1578	6	5	1.200	0.767
bladder	1649	7.41	0.94	1.00	2203	6	5	1.200	0.642
blossom	890	6.79	0.94	1.00	1177	7	6	1.167	0.781
blubber	140	4.94	0.97	0.96	1556	5	5	1.000	0.649
boycott	3052	8.02	1.00	1.00	839	6	5	1.200	0.801
brittle	737	6.60	0.97	1.00	1689	6	6	1.000	0.517
buffalo	8343	9.03	0.97	1.00	1001	7	6	1.167	0.741
cabbage	1124	7.03	0.97	1.00	959	5	5	1.000	0.728
channel	40054	10.60	0.97	1.00	1753	6	5	1.200	0.692
chatter	874	6.77	0.97	1.00	3012	5	4	1.250	0.691
cheddar	427	6.06	0.94	0.96	1909	5	4	1.250	0.587
chipper	483	6.18	0.97	1.00	1910	4	4	1.000	0.621
clatter	98	4.59	0.88	1.00	3001	6	5	1.200	0.699
clobber	418	6.04	0.88	1.00	1600	6	5	1.200	0.657
clutter	1435	7.27	0.97	0.96	2222	6	5	1.200	0.690
collect	14307	9.57	0.97	1.00	1874	7	6	1.167	0.786
college	77818	11.26	1.00	0.96	1742	6	5	1.200	0.827
collide	670	6.51	0.94	1.00	1931	6	5	1.200	0.751
comment	41761	10.64	1.00	0.96	2108	7	6	1.167	0.786
compass	1898	7.55	0.94	0.93	1448	6	6	1.000	0.783
confess	2702	7.90	1.00	1.00	2341	7	6	1.167	0.890

Table A1 Continued

Word	Freq	Log Freq	Acc. LD	Acc. Nmg	BG Freq	UP	*N* Phon.	UP / *N* Phon.	Duration
connect	37477	10.53	1.00	0.96	2027	7	6	1.167	0.667
correct	77798	11.26	0.97	1.00	2131	6	5	1.200	0.785
corrupt	6056	8.71	0.94	1.00	1206	6	5	1.200	0.663
cottage	1318	7.18	1.00	1.00	1320	5	5	1.000	0.741
cripple	929	6.83	0.94	1.00	1483	7	6	1.167	0.715
critter	842	6.74	1.00	1.00	2846	6	5	1.200	0.655
cunning	1063	6.97	0.97	1.00	2535	6	5	1.200	0.615
discuss	30217	10.32	1.00	1.00	1535	7	6	1.167	0.843
dribble	799	6.68	0.91	1.00	1406	7	6	1.167	0.560
drizzle	283	5.65	0.91	1.00	1233	7	6	1.167	0.673
embassy	4493	8.41	0.94	0.96	911	7	6	1.167	0.779
essence	7349	8.90	0.91	1.00	2490	5	5	1.000	0.758
flannel	448	6.11	0.97	0.96	1725	7	6	1.167	0.750
flatter	723	6.58	0.97	0.96	2984	6	5	1.200	0.730
flutter	334	5.81	0.97	1.00	2204	6	5	1.200	0.714
gallery	4490	8.41	0.97	1.00	2585	6	5	1.200	0.669
gimmick	1439	7.27	0.88	1.00	1075	6	5	1.200	0.600
giraffe	379	5.94	0.94	0.96	951	5	4	1.250	0.772
glimmer	401	5.99	0.94	1.00	2059	6	5	1.200	0.617
glitter	898	6.80	0.91	1.00	2765	6	5	1.200	0.640
gorilla	1346	7.21	1.00	1.00	1843	6	5	1.200	0.687
hammock	167	5.12	0.97	1.00	838	6	5	1.200	0.690
harness	2870	7.96	0.82	1.00	2153	7	6	1.167	0.759
immense	2256	7.72	0.91	1.00	1788	6	5	1.200	0.770
lettuce	970	6.88	1.00	0.93	1376	5	5	1.000	0.700
lottery	5648	8.64	1.00	1.00	2373	6	5	1.200	0.726
luggage	1659	7.41	0.97	1.00	624	4	5	0.800	0.763
lullaby	260	5.56	0.88	1.00	1055	7	6	1.167	0.906
mammoth	2688	7.90	0.94	1.00	1033	6	5	1.200	0.642
massage	4037	8.30	0.91	1.00	1239	6	5	1.200	0.542
message	247136	12.42	0.97	0.96	1755	6	5	1.200	0.802
million	55862	10.93	1.00	1.00	2381	7	6	1.167	0.733
missile	5867	8.68	0.97	1.00	2041	6	5	1.200	0.627
mission	28847	10.27	0.97	1.00	2419	6	5	1.200	0.660
narrate	70	4.25	0.94	0.92	2776	6	5	1.200	0.758
offense	8046	8.99	1.00	0.92	1480	6	5	1.200	0.770
pattern	19583	9.88	1.00	1.00	2830	6	5	1.200	0.686

Table A1 Continued

Word	Freq	Log Freq	Acc. LD	Acc. Nmg	BG Freq	UP	*N* Phon.	UP / *N* Phon.	Duration
platter	936	6.84	0.97	1.00	3034	6	5	1.200	0.677
pollute	628	6.44	1.00	1.00	1651	6	5	1.200	0.688
profess	766	6.64	0.91	1.00	1720	7	6	1.167	0.855
quarrel	631	6.45	0.94	1.00	1800	7	6	1.167	0.685
rubbish	7126	8.87	0.91	1.00	981	5	5	1.000	0.716
scatter	845	6.74	0.97	1.00	3015	6	5	1.200	0.718
scuffle	128	4.85	0.88	1.00	992	7	6	1.167	0.760
session	27845	10.23	1.00	1.00	2817	6	5	1.200	0.748
shallow	3761	8.23	1.00	1.00	1546	5	4	1.250	0.771
shatter	2382	7.78	0.94	1.00	2975	5	4	1.250	0.640
sheriff	1740	7.46	0.97	1.00	2164	6	5	1.200	0.691
shuffle	1959	7.58	1.00	1.00	1000	6	5	1.200	0.726
shuttle	8379	9.03	0.94	1.00	1173	6	5	1.200	0.669
skillet	654	6.48	0.94	1.00	1412	6	6	1.000	0.753
skipper	586	6.37	1.00	1.00	1585	6	5	1.200	0.688
slobber	125	4.83	0.94	1.00	1576	6	5	1.200	0.743
smuggle	335	5.81	0.97	1.00	882	7	6	1.167	0.771
sniffle	148	5.00	0.97	1.00	1080	7	6	1.167	0.730
snuggle	278	5.63	1.00	0.96	808	5	6	0.833	0.796
sparrow	710	6.57	0.97	1.00	1383	6	5	1.200	0.831
sputter	198	5.29	0.85	1.00	2233	6	5	1.200	0.748
stagger	353	5.87	0.97	0.96	2306	6	5	1.200	0.714
stammer	103	4.64	0.85	1.00	2445	5	5	1.000	0.731
stirrup	166	5.11	0.91	0.96	1800	6	5	1.200	0.720
stubble	259	5.56	0.88	1.00	1575	7	6	1.167	0.732
suggest	50721	10.83	0.97	1.00	1780	7	6	1.167	0.861
support	248197	12.42	1.00	1.00	1101	7	6	1.167	0.781
swagger	64	4.16	0.82	1.00	1543	6	5	1.200	0.805
swallow	3184	8.07	0.97	1.00	1283	6	5	1.200	0.784
terrace	1715	7.45	0.97	1.00	2890	6	5	1.200	0.741
terrain	6148	8.72	0.91	0.96	3677	5	4	1.250	0.748
tobacco	5750	8.66	1.00	1.00	1111	7	6	1.167	0.845
traffic	25455	10.15	0.97	1.00	1421	7	6	1.167	0.765
trigger	7624	8.94	1.00	1.00	2165	6	5	1.200	0.665
trolley	378	5.94	0.97	1.00	1733	6	5	1.200	0.736
truffle	110	4.70	0.91	1.00	1128	7	6	1.167	0.715
twiddle	4721	8.46	0.82	0.96	893	5	6	0.833	0.632
vaccine	6030	8.71	0.97	0.96	2054	7	6	1.167	0.884

Table A1 Continued

Word	Freq	Log Freq	Acc. LD	Acc. Nmg	BG Freq	UP	*N* Phon.	UP / *N* Phon.	Duration
vanilla	3416	8.14	0.94	0.96	1758	6	6	1.000	0.640
village	13243	9.49	1.00	0.93	1307	6	5	1.200	0.713
villain	2550	7.84	0.91	1.00	2277	6	5	1.200	0.635
warrant	6476	8.78	0.97	1.00	2303	7	6	1.167	0.776
witness	8308	9.03	1.00	1.00	1841	7	6	1.167	0.722

[i] Raw frequency in HAL corpus (English Lexicon Project, Balota et al., 2007)
[ii] Log-frequency in HAL corpus (English Lexicon Project, Balota et al., 2007)
[iii] Accuracy rate in lexical decision (English Lexicon Project, Balota et al., 2007)
[iv] Accuracy rate in spoken word naming from written cue (English Lexicon Project, Balota et al., 2007)
[v] Mean bigram frequency (English Lexicon Project, Balota et al., 2007)
[vi] Phonological uniqueness point (CMU Pronouncing Dictionary, 2015)
[vii] Number of phonemes (CMU Pronouncing Dictionary, 2015)
[viii] Duration of the acoustic stimuli used

Table A2
Characteristics of non-geminate stimuli words

Word	Freq	Log Freq	Acc. LD	Acc. Nmg	BG Freq	UP	*N* Phon.	UP / *N* Phon.	Duration
abdomen	1279	7.15	0.88	0.96	1397	4	7	0.571	0.773
achieve	15544	9.65	0.97	1.00	1284	5	4	1.250	0.754
acrobat	2449	7.80	0.97	1.00	1620	8	7	1.143	0.871
another	286263	12.57	1.00	1.00	2404	6	5	1.200	0.674
auction	70714	11.17	0.91	1.00	2324	6	5	1.200	0.767
average	61913	11.03	1.00	1.00	2259	5	5	1.000	0.769
beneath	5784	8.66	0.97	1.00	2378	5	5	1.000	0.768
bolster	614	6.42	0.91	0.96	2767	7	6	1.167	0.733
caliber	2112	7.66	0.91	1.00	2505	6	6	1.000	0.757
capsize	35	3.56	0.91	1.00	963	7	6	1.167	0.920
caption	1628	7.40	0.94	0.92	2488	7	6	1.167	0.753
carnage	1544	7.34	0.85	1.00	1444	6	6	1.000	0.739
ceramic	1792	7.49	0.88	1.00	2586	7	6	1.167	0.805
chamber	6708	8.81	1.00	1.00	1843	6	5	1.200	0.691
chapter	22942	10.04	0.94	1.00	2409	6	5	1.200	0.697
chariot	878	6.78	0.97	1.00	1986	7	6	1.167	0.804
chimney	536	6.28	0.94	1.00	1010	6	5	1.200	0.674
chowder	199	5.29	0.91	1.00	2032	4	4	1.000	0.703
cluster	7044	8.86	1.00	1.00	2792	7	6	1.167	0.753
comfort	6490	8.78	1.00	1.00	1343	7	6	1.167	0.767
compare	21932	10.00	1.00	1.00	2196	7	6	1.167	0.820
compile	16873	9.73	0.97	1.00	1706	7	6	1.167	0.866
concise	2776	7.93	0.91	1.00	2376	7	6	1.167	0.900
confide	147	4.99	0.88	0.96	1944	7	6	1.167	0.892
console	10876	9.29	0.97	0.93	2405	5	6	0.833	0.842
consult	7290	8.89	0.91	1.00	1867	8	7	1.143	0.870
contact	141845	11.86	1.00	1.00	2430	8	7	1.143	0.957
contend	2083	7.64	0.97	1.00	3345	8	7	1.143	0.867
content	33314	10.41	1.00	1.00	3554	8	7	1.143	0.848
contour	1656	7.41	0.91	1.00	2415	7	6	1.167	0.778
convict	878	6.78	0.97	1.00	1918	8	7	1.143	0.758
costume	3384	8.13	0.97	1.00	1666	7	6	1.167	0.842
country	106711	11.58	1.00	1.00	1891	7	6	1.167	0.752
crystal	13437	9.51	0.91	1.00	1695	8	7	1.143	0.718
cubicle	286	5.66	0.94	0.96	1323	9	8	1.125	0.846
curtain	1953	7.58	0.97	0.92	2118	6	5	1.200	0.663

Table A2 Continued

Word	Freq[a]	Log Freq[b]	Acc. LD[c]	Acc. Nmg[d]	BG Freq[e]	UP[f]	*N* Phon.[g]	UP / *N* Phon.	Duration[h]
custard	519	6.25	0.94	1.00	1860	6	6	1.000	0.759
custody	4502	8.41	1.00	1.00	1290	6	7	0.857	0.784
degrade	1194	7.09	0.82	1.00	1758	7	6	1.167	0.733
detract	717	6.58	0.88	0.96	1893	8	7	1.143	0.826
develop	35652	10.48	0.94	1.00	1489	8	7	1.143	0.795
diploma	1486	7.30	0.97	1.00	1258	7	7	1.000	0.792
element	15056	9.62	0.94	1.00	2456	8	7	1.143	0.668
exhaust	4006	8.30	0.97	1.00	1215	7	6	1.167	0.861
extreme	15968	9.68	1.00	1.00	1618	8	7	1.143	0.842
figment	552	6.31	0.81	0.92	1709	8	7	1.143	0.751
flicker	1600	7.38	0.91	1.00	2253	6	5	1.200	0.703
forgive	8912	9.10	1.00	1.00	1167	6	5	1.200	0.698
fortune	15978	9.68	1.00	1.00	1536	7	6	1.167	0.820
fragile	2339	7.76	1.00	1.00	1616	4	6	0.667	0.827
garbage	12598	9.44	1.00	1.00	1102	6	6	1.000	0.739
grumble	464	6.14	0.94	1.00	1178	8	7	1.143	0.673
hamster	923	6.83	1.00	1.00	2735	7	6	1.167	0.738
hormone	1763	7.48	0.94	1.00	2063	7	6	1.167	0.759
hostage	1293	7.17	0.97	1.00	1579	6	6	1.000	0.903
impeach	198	5.29	0.97	1.00	1373	6	5	1.200	0.801
instead	106318	11.57	0.94	1.00	3265	5	6	0.833	0.736
isolate	3142	8.05	1.00	1.00	2575	7	6	1.167	0.868
janitor	573	6.35	1.00	1.00	1903	7	6	1.167	0.722
knuckle	664	6.50	0.91	1.00	886	6	5	1.200	0.647
lantern	2064	7.63	0.97	1.00	3258	7	6	1.167	0.798
measure	19777	9.89	0.97	1.00	1961	5	4	1.250	0.637
monitor	45909	10.73	1.00	1.00	2231	7	6	1.167	0.663
monster	15607	9.66	1.00	1.00	3591	7	6	1.167	0.745
neglect	1923	7.56	1.00	0.96	1500	8	7	1.143	0.809
nourish	192	5.26	0.94	1.00	1793	5	4	1.250	0.801
nurture	961	6.87	0.88	1.00	1544	5	4	1.250	0.771
oblique[a]	455	6.12	0.94	1.00	878	6	5	1.200	0.705
obscure	5753	8.66	1.00	1.00	1336	8	7	1.143	0.860
parsley	784	6.66	0.97	1.00	1602	6	6	1.000	0.677
partner	17565	9.77	0.97	1.00	2261	7	6	1.167	0.623
perfume	1151	7.05	0.94	1.00	1790	7	6	1.167	0.789

Table A2 Continued

Word	Freq[a]	Log Freq[b]	Acc. LD[c]	Acc. Nmg[d]	BG Freq[e]	UP[f]	*N* Phon.[g]	UP / *N* Phon.	Duration[h]
plaster	737	6.60	0.85	0.97	3062	7	6	1.167	0.713
posture	1909	7.55	1.00	1.00	2020	6	5	1.200	0.737
poverty	7521	8.93	0.97	1.00	1897	4	6	0.667	0.726
primate	2053	7.63	1.00	0.96	2560	7	6	1.167	0.682
profile	11524	9.35	1.00	1.00	1556	7	6	1.167	0.816
prosper	1652	7.41	1.00	1.00	2164	7	6	1.167	0.801
protect	29349	10.29	0.88	1.00	1871	8	7	1.143	0.839
puberty	1014	6.92	0.97	1.00	1583	5	7	0.714	0.737
publish	10307	9.24	0.97	1.00	1459	7	6	1.167	0.783
quantum	16245	9.70	0.84	1.00	1475	8	7	1.143	0.765
realize	34837	10.46	0.97	1.00	2212	7	6	1.167	0.936
rejoice	878	6.78	0.97	1.00	1500	6	5	1.200	0.880
relieve	2005	7.60	0.94	0.96	2110	6	5	1.200	0.744
remorse	736	6.60	0.94	0.96	2129	7	6	1.167	0.857
reserve	26384	10.18	0.94	1.00	3224	6	5	1.200	0.757
respect	37713	10.54	0.97	1.00	2327	8	7	1.143	0.836
sausage	1273	7.15	0.97	1.00	901	5	5	1.000	0.881
scratch	8067	9.00	1.00	1.00	1787	6	5	1.200	0.856
serious	55817	10.93	1.00	1.00	2786	7	6	1.167	0.853
several	170923	12.05	1.00	1.00	2791	5	6	0.833	0.724
shrivel	137	4.92	0.91	1.00	1488	7	6	1.167	0.743
sincere	4659	8.45	0.94	0.96	3721	7	6	1.167	0.841
slander	2098	7.65	0.94	1.00	2743	7	6	1.167	0.782
slender	1810	7.50	0.90	1.00	3031	5	6	0.833	0.742
slumber	344	5.84	0.91	1.00	1515	7	6	1.167	0.745
soldier	8296	9.02	0.97	0.92	1950	6	5	1.200	0.793
solicit	1219	7.11	0.94	1.00	1714	8	7	1.143	0.816
sparkle	615	6.42	0.97	1.00	1453	8	7	1.143	0.774
special	105765	11.57	1.00	1.00	1494	7	6	1.167	0.750
speckle	98	4.59	0.94	1.00	1382	7	6	1.167	0.768
steward	1151	7.05	0.94	0.96	2088	6	5	1.200	0.848
sublime	885	6.79	0.94	0.96	1299	5	6	0.833	0.884
supreme	16923	9.74	0.94	1.00	1651	5	6	0.833	0.811
swindle	242	5.49	0.97	1.00	2175	8	7	1.143	0.694
symptom	2367	7.77	0.97	1.00	699	8	7	1.143	0.820
tantrum	639	6.46	0.84	1.00	1893	8	7	1.143	0.770

Table A2 Continued

Word	Freq[a]	Log Freq[b]	Acc. LD[c]	Acc. Nmg[d]	BG Freq[e]	UP[f]	*N* Phon.[g]	UP / *N* Phon.	Duration[h]
texture	5678	8.64	1.00	1.00	1960	7	6	1.167	0.723
treason	1613	7.39	0.88	1.00	2481	7	6	1.167	0.715
version	224831	12.32	1.00	1.00	3059	6	5	1.200	0.703
veteran	7287	8.89	0.94	0.96	3403	7	6	1.167	0.716
vitamin	3789	8.24	0.97	1.00	2341	8	7	1.143	0.741
whisper	1759	7.47	1.00	1.00	2122	6	5	1.200	0.737
whistle	2672	7.89	1.00	1.00	1913	6	5	1.200	0.629