Title
Simulating length and frequency effects across multiple tasks with the Bayesianmodel BRAID-Phon

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 42(0)

Authors
Saghiran, Ali
Valdois, Sylviane
Diard, Julien

Publication Date
2020

Simulating length and frequency effects across multiple tasks with the Bayesian model BRAID-Phon

Ali Saghiran (ali.saghiran@univ-grenoble-alpes.fr)
Univ. Grenoble Alpes, Univ. Savoie Mont Blanc, CNRS, LPNC, 38000 Grenoble, France

Sylviane Valdois (sylviane.valdois@univ-grenoble-alpes.fr)
Univ. Grenoble Alpes, Univ. Savoie Mont Blanc, CNRS, LPNC, 38000 Grenoble, France

Julien Diard (julien.diard@univ-grenoble-alpes.fr)
Univ. Grenoble Alpes, Univ. Savoie Mont Blanc, CNRS, LPNC, 38000 Grenoble, France

Abstract

In visual word processing modeling, few models have successfully accounted for a large variety of tasks, and large corpora of behavioral observations. We consider a dataset from a megastudy, in which participants performed three tasks (lexical decision, word naming, and word recognition in a progressive demasking situation), on the same, large set of stimuli. We define the BRAID-Phon model, an extension of a previous probabilistic model, the BRAID model, whose originality is its visuo-attentional component, in which a visuo-attentional distribution spatially deploys sensory processing capabilities. BRAID-Phon includes phonological representations of words, allowing simulating the naming task. We simulated the three tasks on the dataset we considered, and analyzed predicted reaction times in terms of word frequency and word length effects. Simulation results show that BRAID-Phon successfully captures the direction and order of magnitude of the observed effects, in all three tasks.

Keywords: Visual word processing; computational modeling; reading aloud; lexical decision; megastudy simulation

Introduction

Different experimental paradigms can be used to study visual word processing through the analysis of participants' response times (RTs). The lexical decision (LD) and word naming (NMG) paradigms are the most popular. In LD, participants have to decide whether an input letter-string is a known word or not; RTs correspond to time needed to press a YES or NO response key. In naming, participants have to read the word aloud and identification RTs are collected with a microphone. Another very useful but less frequently used visual word processing paradigm is the progressive-demasking (PDM) task. In this task, a word is progressively revealed and participants press the key as soon as they think they have identified it. Response accuracy is checked after their response, by asking them to spell the word they have identified.

Using these different paradigms, a considerable amount of data has been accumulated that led to identifying the most relevant psycholinguistic factors that influence the reader's performance. When the same set of words was used in the different tasks, allowing cross-task comparisons, results revealed that response times were highly inter-correlated, suggesting that the different tasks involve shared cognitive processes (Carreiras, Perea, & Grainger, 1997). However, the current literature lacks an integrated model of visual word processing, able to account for the variety of observations across different cognitive tasks (Norris, 2013).

In the last decade, the *megastudy* approach has become increasingly popular. Megastudies provide experimental data for thousands of items, over tens or hundreds of participants. Even if concerns have been raised about the relative stability and reliability of the collected datasets (Sibley, Kello, & Seidenberg, 2009), such megastudies minimize bias risks from working with short lists of stimuli and help to reduce inconsistencies in the word recognition literature (Balota, Yap, Hutchison, & Cortese, 2012). However, most of these large data sets were collected on a single task, mainly LD; only some data were collected in both LD and NMG, and to the best of our knowledge, a single megastudy (the Chronolex database) provides data for the same large set of words in the three tasks of LD, NMG and PDM (Ferrand et al., 2011). Thus, the Chronolex dataset is particularly appropriate to explore the word processing effects that are common to the different tasks and how these effects are modulated depending on task demands.

Different classes of computational models have been developed that implement different aspects of the reading process. Word recognition models focus on the visual processes involved in letter identification and how perceptual information contacts orthographic knowledge during processing (see Norris (2013), for a review). They are mainly used to simulate the behavioral effects reported in lexical decision. Models of naming more specifically question how a spoken form of the word is generated from the printed letter string (Ans, Carbonnel, & Valdois, 1998; Coltheart, Rastle, Perry, Langdon, & Ziegler, 2001; Perry, Ziegler, & Zorzi, 2007, 2010). They are typically used to account for RTs in reading aloud. Naming models can further account for LD and provide a reasonable fit to human data on this task. However, most of these models make simplifying assumptions about the visual word recognition mechanisms and/or letter position encoding, which limits their explanatory power. Moreover, only a few computational models have been confronted to large-scale datasets, and when they were, simulations were limited to a single task (Norris & Kinoshita, 2012; Perry, Ziegler, & Zorzi, 2014; Sibley, Kello, & Seidenberg, 2010). The availability of cross-task megastudies (Balota et al., 2007; Ferrand et al., 2011) opens new perspectives, allowing to check the models' ability to account for robust effects reported in different experimental paradigms.

The originality of the current study is twofold. First, our purpose is to develop a new computational model of the reading process that includes both a fully specified word recognition component and a phonological component, thus being able to simulate relevant behavioral effects from a variety of tasks. Second, we ask how well the model can account for cross-task large-scale datasets, focusing on the two main effects of word frequency and word length which are known as the best predictors of RTs in LD, NMG and PDM (Balota, Cortese, Sergent-Marshall, Spieler, & Yap, 2004; Ferrand et al., 2011).

For this purpose, we start from the BRAID model ("Bayesian word Recognition with Attention, Interference and Dynamics", Phénix (2018)) and extend it by adding a phonological component. BRAID is the first word recognition model that implements a variety of visual processes known to affect letter-string processing, namely an acuity gradient, lateral interference between adjacent letters and visual attention deployment. The model was previously shown to account for classical effects in letter perception, word recognition and lexical decision, such as frequency effects, word superiority effects, neighborhood effects, transposed-letter priming effects or the optimal viewing position effect (Phénix, 2018; Phénix, Valdois, & Diard, 2018; Phénix, Ginestet, Valdois, & Diard, submitted). Further, BRAID was successful at simulating length effects in LD as reported in the French Lexical Project megastudy (Ginestet, Phénix, Diard, & Valdois, 2019).

In the next section, we describe how BRAID was extended into "BRAID-Phon", a "BRAID with phonology" model, to develop the first model of reading aloud that incorporates a fully specified word recognition sub-model. Second, we show how BRAID-Phon simulates the LD, NMG and PDM tasks, with a self-imposed constraint to simulate all three tasks using the same default values of the parameters. We also describe the original dataset from the Chronolex megastudy (Ferrand et al., 2011). Finally, we analyse and discuss experimental results from our simulations, by focusing on the frequency and length effects, that, in the Chronolex data, affect human RTs in the same direction in all three tasks but with different magnitudes.

Model

BRAID

BRAID is a computational model of the visual processes and knowledge involved in letter recognition, word recognition and lexical decision. Probabilistic variables and probability distributions model representations and knowledge, and probabilistic inference provides mathematical expressions for simulated tasks. The BRAID model features the three classical levels of visual processing, as in the seminal IA model (McClelland & Rumelhart, 1981): the letter sensory level, the letter perceptual level and the lexical level (see Fig 1). The main originality of BRAID is the inclusion of a visual attention layer, between the sensory layer and the letter perception layer. Here, we only provide a rapid description of each submodel; the full mathematical description is available elsewhere (Phénix, 2018).

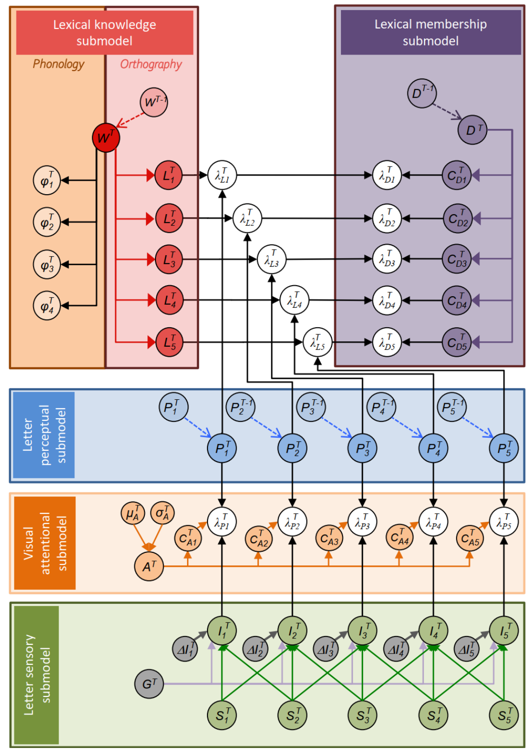

Figure 1: Graphical representation of the BRAID and BRAID-Phon models. Each submodel is represented by a colored, labelled rectangle. Variables are represented by nodes and arrows refer to probabilistic dependencies. The full structure is the one of the BRAID-Phon model; removing the "phonology" portion of the lexical submodel yields the structure of the initial BRAID model.

The *sensory submodel* of BRAID implements low-level visual processing, taking into account effects from an acuity gradient, from lateral interference between letters of the stimulus, and from visual similarity between letters. S^t_n variables describe the stimulus at each time step t and each position n, G^t represent the gaze position at time step t. Letter processing is parallel and results in probability distributions over variables I^t_n, that represent internal letter representation.

The *visual-attention submodel* spatially modulates the transfer of information propagation during sensory processing, from the sensory to the perceptual submodels, to favor processing of the attended portion of the stimulus, to the detriment of the rest. In other words, it can be construed as a spatial filter of information across letter positions. The distribution of attention is represented by a probability distribution, assumed to be Normal over spatial positions, so that it is described by its location μ^t_A and dispersion σ^t_A.

The *letter perceptual submodel* is a dynamic model that implements accumulation of perceptual evidence. Probability distributions over variables P^t_n evolve over time as they receive information from sensory processing, allowing to identify letters at each position n and each time step t.

The *lexical knowledge submodel* represents knowledge about the orthography of known words, with a naïve Bayes model predicting, for each word W, its letter L_n^t at each position. In the present study, the lexical submodel is configured with the *Lexique 3.82* database, featuring the spelling of 92,117 French words (New, 2006). The lexical submodel also includes a dynamical model, to accumulate perceptual evidence about which known word corresponds to the stimulus, if one does. The initial state of this dynamical model is the prior distribution $P(W^0)$, which is identified to word frequency from the reference lexicon.

The final submodel, the *lexical membership submodel*, allows evaluating the correspondence between the stimulus and known words, using a dynamically evolving probability distribution over Boolean variable D^t. In other words, this submodel implements an "error model": assuming the input string is a word ($D^t = True$), perceived letters should match those of a known word in all positions; on the contrary, if the input is not a word ($D^t = False$), matching should fail in at least one position.

In the overall architecture of the BRAID model, information propagates at each time step both in a bottom-up manner (from the sensory to the perceptual to both the lexical knowledge and lexical membership submodels) and in a top-down manner (from the lexical submodels to the perceptual submodel). Therefore, lexical top-down information influences processing at the letter level, yielding word and pseudo-word superiority effects, as in classical models. Further, this influence is modulated by lexical membership: if the stimulus is probably not a word, top-down influence is decreased so that only perceptual, bottom-up information contributes to letter identification probabilities.

BRAID-Phon

In order to extend the BRAID word recognition model into the BRAID-Phon reading aloud model, we add a phonological sub-layer, extending the lexical knowledge submodel (see Fig 1). We represent the phonological description of each known word in a mirror manner to orthography: the probability distribution over phonemes $\varphi_{1:M}^t$, for all positions 1 to M, is defined by a time-invariant naïve Bayes model. M is the maximal phonological length of words in the lexicon and a specific phoneme value (#) is used to indicate the end of the phonological sequence for shorter words.

For simplicity, we only represent words by phonemes at each position, and do not address issues about the nature of phonological representations here (e.g., syllabic vs phonemic). As for orthographic knowledge, in the experiment presented here, we identify phonological knowledge from the *Lexique 3.82* database (New, 2006).

Simulations

Since BRAID-Phon is a probabilistic model, it is formally defined by a joint probability distribution, allowing to formulate and compute all possible conditional probability distributions related to variables of the model. Therefore, simulating a given cognitive task amounts to choosing the variable of interest X^T, describing the input, known values (usually describing the stimulus $S_{1:N}^{1:T}$ at all positions and for all time steps, gaze position $G^{1:T}$ and the parameters $\mu_A^{1:T}$, $\sigma_A^{1:T}$ of the attention distribution), and computing the corresponding probability distribution $P(X^T \mid S_{1:N}^{1:T}\ G^{1:T}\ \mu_A^{1:T}\ \sigma_A^{1:T})$.

Lexical Decision For instance, computing the probability distribution over D^t, the dynamic variable of the lexical membership submodel, allows simulating the lexical decision task (Phénix, 2018; Ginestet et al., 2019). Mathematically:

$$Q_{LD}^T = P\Big(D^T \mid \mu_A^{1:T}\ \sigma_A^{1:T}\ S_{1:N}^{1:T}\ G^{1:T}\ [\lambda_{P}{}_{1:N}^{1:T} = 1]\ [\lambda_{D}{}_{1:N}^{1:T} = 1]\Big)$$

The lambda variables $\lambda_{P}{}_{1:N}^{1:T}$ and $\lambda_{D}{}_{1:N}^{1:T}$ are *coherence* variables that allow selecting how information propagates in the model (Gilet, Diard, & Bessière, 2011). Here, assuming they are set to 1 means that information propagates throughout the whole BRAID-Phon architecture. Computing ${Q_{LD}}^T$ yields the evolution over time of the probability value of the "yes"-answer to the LD task ($D^T = True$) for a given orthographic stimulus $S_{1:N}^{0:T}$.

Naming In a similar manner, the naming task is simulated by computing:

$$Q_{NMG}^T = P\Big(\varphi_{1:M}^T \mid \mu_A^{1:T}\ \sigma_A^{1:T}\ S_{1:N}^{1:T}\ G^{1:T}\ [\lambda_{P}{}_{1:N}^{1:T} = 1]\ [\lambda_{D}{}_{1:N}^{1:T} = 1]\Big)$$

Therefore, reading aloud is simulated in *BRAID-Phon* by inferring the probability distributions over phonemes at each time step, given an orthographic input, and their evolution as sensory processing of the visual input proceeds. In simulations, NMG RTs are based on the probability distribution over the first phoneme, provided that the whole sequence of phonemes is correct. These criteria match the task conditions, since behavioral RTs correspond to the beginning of the pronunciation and are only reported for correct responses.

Progressive demasking The third task, progressive demasking, is simulated in BRAID-Phon as word recognition was simulated in BRAID (Phénix, 2018), that is, we compute the probability distribution over words at each time step:

$$Q_{WR}^T = P\big(W^T \mid \mu_A^{1:T}\ \sigma_A^{1:T}\ S_{1:N}^{1:T}\ G^{1:T}\ [\lambda_{P}{}_{1:N}^{1:T} = 1]\big)$$

However, stimulus presentation is particular in PDM, since the word is gradually revealed. To do this, time is considered as a sequence of periods (of 210 ms each in the Chronolex experiment), each period divided further into a sequence of cycles (of 14 ms each, so that there are 15 cycles per period). Each period first shows the stimulus for a portion of the cycle, then a mask made of # characters for the remainder of the cycle. Initially, the stimulus is shown for 1 cycle and the mask for 14 cycles, then 2 and 13, 3 and 12, etc.

This is easily simulated in BRAID-Phon by adequately setting the input values of stimulus variables $S_{1:N}^{1:T}$, alternating between those of an input letter sequence and those of an input mask sequence, as a function of time.

Experimental Data The Chronolex study (Ferrand et al., 2011) provides a database of RTs for 1,482 French words, each corresponding to the mean RT collected over 105 participants. The dataset contains only monosyllabic and monomorphemic words of different lengths, from 2 to 8 letters, with, understandably, few 8-letter monosyllabic words compared to other lengths.

Procedure The BRAID model was previously calibrated (Phénix, 2018) so that one simulated iteration corresponds to one millisecond of physical time. In this work, we use the default values of all parameters. In particular, the gaze position g^T and the position of attentional focus μ_A^T coincide, and are set to the central position of the stimulus ($g^T = \mu_A^T = (N+1)/2$, with N the stimulus length). The standard deviation of the attention distribution is set to $\sigma_A^T = 1.75$.

Once run, simulations yield the evolution over time of probability values of the variable of interest. A straightforward decision model is considered: to obtain simulated RTs, we compute the number of iterations needed for the probability value to reach a fixed decision threshold.

In the three considered tasks, the domains of the variables of interest have different sizes: 2 for the lexical membership variable D^T, 40 for the first phoneme variable φ_1^T and 92,117 for the word variable W^T. Therefore, we allow for three different decision threshold values. To calibrate these (Ginestet et al., 2019), we run simulations for 1,000 iterations for the LD and NMG tasks, and 3,000 for the PDM task, then explore a large set of decision thresholds, yielding error rates and a mean square error to the behavioral data. Decision thresholds are selected to solve the trade-off between these two measures (details not provided here). Therefore, below, we report results for the following threshold values: $\tau_{LD} = 0.9$, $\tau_{NMG} = 0.8$ and $\tau_{PDM} = 0.5$. Items for which the model provides an incorrect output or for which the considered probability curve does not reach the specified threshold are not taken into account in the analyses. We note further that the decision threshold for lexical decision τ_{LD} has the same value as in previous work, which was based on an entirely different dataset (Ginestet et al., 2019).

In order to compare the simulated RTs to the behavioral RTs of the Chronolex study, we perform linear regressions for each task, using two predictors: word frequency and word length. Additionally, since NMG data is collected by automatic analyses of audio recordings, it is affected by voice detection biases, stress patterns, etc. Therefore, the Chronolex NMG task regression also includes, as predictors, 13 control variables related to the articulatory nature of the stimulus onset. Such analyses are common in the field (Spieler & Balota, 1997).

Results

BRAID-Phon correctly recognizes 98% of items in LD, 88.5% in NMG and 90.7% in PDM. Figure 2 and Figure 3 respectively show regression lines for the frequency and length effects, for behavioral and simulated RTs.

In the LD task, the two length and frequency factors explain 39% of simulated data variance ($F(2,1458) = 316.4$; $p < .001$) against 41% in the Chronolex data ($F(2,1458) = 524.6$; $p < .001$). The model provides an accurate simulation of the log-frequency effect: -30 iterations per log(ppm) in simulation versus -38 ms per log(ppm) in the Chronolex data. The length effect is also well predicted: 15 iterations per letter in simulation and 8 ms per letter in the Chronolex data. The corresponding partial R^2 are also similar: 0.20 in simulation versus 0.29 in the behavioral data for the frequency effect and 0.06 versus 0.02 for the length effect.

In the NMG task, frequency and length explain 24% of simulated data variance ($F(2,1391) = 220.4$; $p < .001$). For Chronolex, the two factors and the onset control variables account for 53% of variance ($F(14,1379) = 113$; $p < .001$). The slope of the frequency effect is -27.5 iterations per log(ppm) in simulation versus -7.5 ms per log(ppm) in the Chronolex data. The slope of the length effect is 23.1 iterations per letter versus 4.9 ms per letter in the Chronolex data. For the frequency effect, the partial R^2 are 0.1 in simulation versus .02 in behavioral data, and are respectively 0.1 versus 0.01 for the length effect. The word onset variables alone account for 45% of Chronolex NMG RTs variability.

In the PDM task, the two predictors of length and frequency explain 38.5% of simulated RTs variance ($F(2,1354) = 425.9$; $p < .001$) and 19.3% in the Chronolex data ($F(2,1354) = 163.2$; $p < .001$). The regression yields large slope values for the simulation data: -170.7 iterations per log(ppm) for the frequency effect and 72.53 iteration per letter when dealing with the length effect (-21.5 and 20.3 respectively in behavioural data). For this task, the partial R^2 of the length effect are equal (0.06) in simulated and Chronolex data. The partial R^2 of the frequency effect is 0.22 in simulated RTs versus 0.06 in the Chronolex data.

Overall, for all tasks and effects, the simulation results contains the same effects, and in the same direction, as observed in the Chronolex data. A direct comparison of RTs for the three tasks and the two effects shows that BRAID-Phon successfully captures the general decrease of RTs with increasing log-frequency and the general increase in RTs with increasing length. As in the experimental data, RTs at the intercept are far longer in PDM than in LD or NMG. However, BRAID simulates similar intercepts for LD and NMG while RTs at the intercept are longer for LD than NMG in the human data. For both the frequency and length effects, the slopes of the regressions are quite comparable in LD for the human data and the simulations. However, we observe that BRAID-Phon tends to overestimate the amplitude of the effects in NMG and simulates exaggerated effects in PDM.

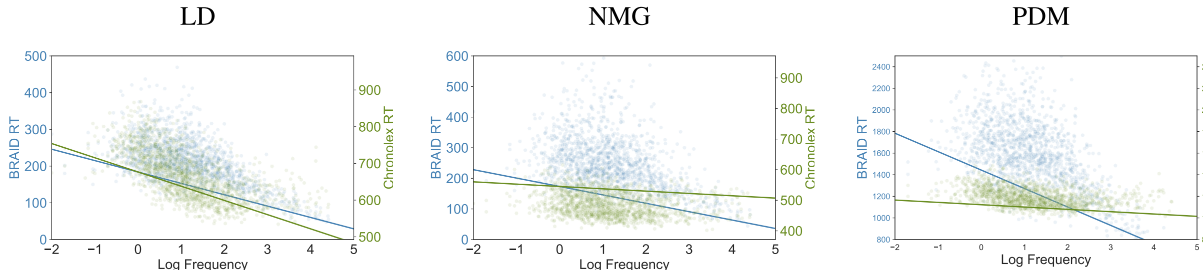

Figure 2: Frequency effect on RTs simulated by BRAID-Phon (blue) and Chronolex RTs (green) and a scatterplot of item-level RTs. To ease the comparison of the slopes, the y-axis of simulated RTs is shifted by a constant equal to the difference of the two regressions intercepts.

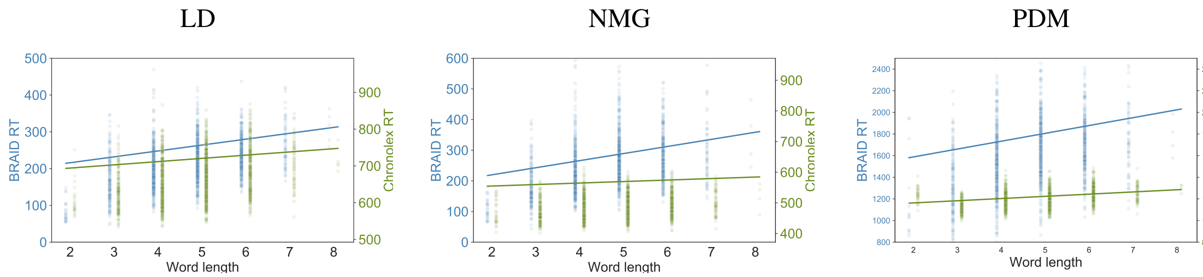

Figure 3: Length effect on RTs simulated by BRAID-Phon (blue) and Chronolex RTs (green) and a scatterplot of item-level RTs. The y-axis of simulated RTs is shifted (see Figure 2).

Discussion

This study is a first attempt to account for the main effects of word frequency and word letter length in three different visual word processing tasks while using the same computational model and the same parameter values. The Bayesian formalism we follow allows to derive from the same model, in a mathematically rigorous manner, the tasks of lexical decision, naming and progressive demasking.

As far as we know, no other computational model has been evaluated on a multiple task study. Several computational models can successfully simulate LD and NMG performance but most of the time evidence comes from very limited datasets. Only one model attempted to simulate the PDM task (see the MROM model, Carreiras et al. (1997)) and the model parameters were adapted to match the experimental conditions. In contrast, the simulations we performed were computed while using the same parameter values for all three tasks. The decision threshold value was taken from previous simulations for LD; thresholds were fixed for the new tasks of NMG and PDM using the same method as previously for LD.

We used BRAID to simulate the very classical effect of frequency that most models have previously simulated. BRAID was further checked for its ability to account for the length effect. Although being a very robust behavioural effect in a variety of tasks, previous models could only account for the length effect in naming. We show that BRAID-Phon successfully predicts the direction and the order of magnitude of the length effect in LD, NMG and PDM. In particular, the model quite successfully simulates length effects in LD, which was challenging for previous models. The model appears to be robust, as similar success was previously reported for BRAID when confronted to the French Lexicon Project dataset (Ferrand et al., 2010) to simulate length effects in LD (Ginestet et al., 2019).

In most theoretical frameworks, the length effect was viewed as inherently related to serial processing (Coltheart et al., 2001; Perry et al., 2007). Counter-intuitively, the length effect in our model is a consequence of the parallel processing of letters and of limited visual attention resources. Given that a limited amount of visual attention is available for letter processing, letters are more efficiently processed in shorter words than in longer words. Therefore, shorter words are recognized faster than longer words. In the model, the same two mechanisms of parallel processing and visual attention limitation account for the deleterious effects of word length on the number of iterations in all three tasks.

Nevertheless, BRAID-Phon is confronted to several limitations in its current implementation. First, the model implements phonological knowledge in a very simplistic way and does not include any information about articulatory features of speech. This may explain some discrepancies between the experimental data and simulated findings. Indeed, behavioral data shows that the identity of the first phoneme explains up to 45% of naming latencies variability (Yap & Balota, 2009), suggesting that, in the NMG task, Rts are affected by articulatory planning and speech production of the word to be named. Such considerations are absent of BRAID-Phon, since it is restricted to stimulus recognition.

Second, BRAID-Phon tends to exacerbate length effects, in particular in PDM. We have previously shown that length effects in LD are better captured when allowing the model to shift attention and gaze position during word processing beyond 6 letter length (Ginestet et al., 2019). Such visuo-attentional shifts would reduce the amplitude of length effects in all three tasks, particularly in PDM, where RTs are around 1.2 s, making visuo-attentional shifts highly likely in participants. Moreover, the PDM task is a particular condition in which participants are instructed to respond as quickly as possible, as soon as they think they have identified the word. The instruction may thus favor guessing strategies which would correspond in the model to higher reliance on lexical knowledge. Better fit are thus expected if we allowed higher top-down influence from the lexical submodel to the letter perception submodel.

References

Ans, B., Carbonnel, S., & Valdois, S. (1998). A connectionist multiple-trace memory model for polysyllabic word reading. *Psychological Review*, *105*(4), 678–723.

Balota, D. A., Cortese, M. J., Sergent-Marshall, S. D., Spieler, D. H., & Yap, M. J. (2004). Visual word recognition of single-syllable words. *Journal of Experimental Psychology: General*, *133*(2), 283–316.

Balota, D. A., Yap, M. J., Hutchison, K. A., & Cortese, M. J. (2012). Megastudies: What do millions (or so) of trials tell us about lexical processing? In J. S. Adelman (Ed.), *Visual word recognition: models and methods, orthography and phonology.* New York: Psychology Press.

Balota, D. A., Yap, M. J., Hutchison, K. A., Cortese, M. J., Kessler, B., Loftis, B., ... Treiman, R. (2007, 01). The English lexicon project. *Behavior Research Methods*, *39*(3), 445–459.

Carreiras, M., Perea, M., & Grainger, J. (1997). Effects of the orthographic neighborhood in visual word recognition: Cross-task comparisons. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *23*(4), 857–871.

Coltheart, M., Rastle, K., Perry, C., Langdon, R., & Ziegler, J. (2001). DRC: A dual route cascaded model of visual word recognition and reading aloud. *Psychological Review*, *108*(1), 204–256.

Ferrand, L., Brysbaert, M., Keuleers, E., New, B., Bonin, P., Méot, A., ... Pallier, C. (2011). Comparing word processing times in naming, lexical decision, and progressive demasking: Evidence from Chronolex. *Frontiers in Psychology*, *2*, 306.

Ferrand, L., New, B., Brysbaert, M., Keuleers, E., Bonin, P., Méot, A., ... Pallier, C. (2010). The French lexicon project: Lexical decision data for 38,840 French words and 38,840 pseudowords. *Behavior Research Methods*, *42*(2), 488–496.

Gilet, E., Diard, J., & Bessière, P. (2011). Bayesian action-perception computational model: Interaction of production and recognition of cursive letters. *PLoS ONE*, *6*(6), e20387.

Ginestet, E., Phénix, T., Diard, J., & Valdois, S. (2019). Modeling the length effect for words in lexical decision: The role of visual attention. *Vision Research*, *159*, 10–20.

McClelland, J. L., & Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception: Part 1. an account of basic findings. *Psychological Review*, *88*(5), 375–407.

New, B. (2006). Lexique 3: Une nouvelle base de données lexicales. In *Actes de la conférence traitement automatique des langues naturelles (TALN 2006).*

Norris, D. (2013). Models of visual word recognition. *Trends in Cognitive Sciences*, *17*(10), 517–524.

Norris, D., & Kinoshita, S. (2012). Reading through a noisy channel: Why there's nothing special about the perception of orthography. *Psychological Review*, *119*(3), 517–545.

Perry, C., Ziegler, J. C., & Zorzi, M. (2007). Nested incremental modeling in the development of computational theories: The CDP+ model of reading aloud. *Psychological Review*, *114*(2), 273–315.

Perry, C., Ziegler, J. C., & Zorzi, M. (2010). Beyond single syllables: Large-scale modeling of reading aloud with the connectionist dual process (CDP++) model. *Cognitive Psychology*, *61*(2), 106–151.

Perry, C., Ziegler, J. C., & Zorzi, M. (2014). When silent letters say more than a thousand words: An implementation and evaluation of CDP++ in French. *Journal of Memory and Language*, *72*, 98–115.

Phénix, T., Ginestet, E., Valdois, S., & Diard, J. (submitted). The role of attention in visual word recognition: A Bayesian modeling approach. –.

Phénix, T., Valdois, S., & Diard, J. (2018). Reconciling opposite neighborhood frequency effects in lexical decision: Evidence from a novel probabilistic model of visual word recognition. In *Proc. of the 40th annual conference of the cognitive science society* (pp. 2238–2243). Madison, WI.

Phénix, T. (2018). *Modélisation bayésienne algorithmique de la reconnaissance visuelle de mots et de l'attention visuelle*. Unpublished doctoral dissertation, Université Grenoble Alpes.

Sibley, D. E., Kello, C. T., & Seidenberg, M. S. (2009). Error, error everywhere: A look at megastudies of word reading. In *Proceedings of the 21th annual conference of the cognitive science society* (pp. 1036–1041). Amsterdam.

Sibley, D. E., Kello, C. T., & Seidenberg, M. S. (2010). Learning orthographic and phonological representations in models of monosyllabic and bisyllabic naming. *European Journal of Cognitive Psychology*, *22*(5), 650–668.

Spieler, D. H., & Balota, D. A. (1997). Bringing computational models of word naming down to the item level. *Psychological Science*, *8*(6), 411–416.

Yap, M. J., & Balota, D. A. (2009). Visual word recognition of multisyllabic words. *Journal of Memory and Language*, *60*(4), 502–529.