Cognitive Systems Research 88 (2024) 101284

Contents lists available at ScienceDirect

Cognitive Systems Research

journal homepage: www.elsevier.com/locate/cogsys

A design for neural network model of continuous reading

Jarkko Hautala [a,d,*,1], Mirka Saarela [b], Otto Loberg [c], Tommi Kärkkäinen [b]

[a] Niilo Mäki Institute, Jyväskylä, Finland
[b] Faculty of Information Technology, University of Jyväskylä, PO Box 35 FI-40014, Finland
[c] Department of Psychology, Bournemouth University, Talbot Campus, Fern Barrow, Poole BH12 5BB, United Kingdom
[d] Centre for Interdisciplinary Brain Research, University of Jyväskylä, Jyväskylä, Finland

ARTICLE INFO

Action editor: S Thill

Keywords:
Eye movement control of reading
Word recognition
Computational modeling
Neural network

ABSTRACT

Cognition and learning are exceedingly modeled as an associative activity of connectionist neural networks. However, only a few such models exist for continuous reading, which involves the delicate coordination of word recognition and eye movements. Moreover, these models are limited to only orthographic level of word processing with predetermined lexicons. Here, we present a conceptual design of a developmentally plausible neural network model of reading designed to simulate word learning, parafoveal preview activation of words, their later foveal word recognition including phonological decoding, and forward saccade length as a control mechanism for intake of new textual information. We will discuss the theoretical advancements of the design and avenues for future developments.

1. A brief historical overview of computational modeling of reading

Reading has attracted computational scientists since Morton's (1969) *logogen* model of word identification. Due to the complexity of reading cognition, scholars have mainly focused on modeling separately different reading subprocesses, particularly visual word recognition of single words at the level of representational activity and eye movement control during continuous reading (Reichle, 2021). Since word recognition is widely believed to be the driving force of guiding eye movements, this segregation of research fields is not optimal (Grainger, 2003). The first attempts to computationally integrate these models have recently been published (Snell et al., 2018; Li & Pollatsek, 2020). We will first provide a brief historical perspective on the previous research leading to these integrative models.

Word recognition models. One of the principal challenges of reading research has been to credibly explain how people can effortlessly recognize a word as a single perception without confusing it to other resembling words. In their seminal work, McClelland and Rumelhart (1981) presented the interactive-activation model (IAM), in which each activated letter node further activated word nodes that contained this letter and inhibited the word nodes that did not contain this letter. Word frequency was operationalized as the resting state activation level of the word nodes. The IAM then provided a basis for developing more comprehensive models of visual word recognition, including mapping from orthography to phonology.[2] To accommodate the findings of readers showing also serial effects (such as length effect) when reading novel words (e.g., Weekes, 1997), this dual-route cascaded model of reading aloud (DRC; Coltheart et al., 2001) included a devoted phonological decoding route, which converts graphemes serially into phonemes according to pronunciation rules to enable reading words not included in its orthographic lexicon. Connectionist dual-process model (CDP+; Perry et al., 2007) then provided a neural network implementation of the decoding route. More recent theoretical advancements that have been computationally implemented include flexible letter encoding to simulate letter substitution and transposition effects (Dandurand et al., 2013; Dandurand et al., 2010; Davis, 2010), visual acuity and crowding influences on letter encoding (Snell et al., 2018), and multisyllabic word reading (Davis, 1999) with phonological syllable stress assignment (CDP++ model by Perry et al., 2010). There have also been important improvements in developing single-route neural network models capable of simulating a wide range of word recognition

* Corresponding author at: Niilo Mäki Foundation, B.O. Box 29 FI-40101, Jyväskylä, Finland.
E-mail address: jarkko.v.hautala@jyu.fi (J. Hautala).
[1] Street address: Asemakatu 4a, 3rd floor, 40100 Jyväskylä, Finland.
[2] Mapping to semantics is left out from the present discussion because mapping from orthography to phonology is not yet addressed by the current integrative models.

https://doi.org/10.1016/j.cogsys.2024.101284
Received 13 March 2023; Received in revised form 18 June 2024; Accepted 5 September 2024
Available online 7 September 2024

phenomena (Sibley et al., 2012). It can be summarized that the visual encoding of letters to the activation of sublexical units and words, and further to phonology, is well scrutinized, albeit the exact neural network design and computational implementation are still works in progress (Dandurand et al., 2013).

Models of eye movement control in reading. Meanwhile, other researchers have been interested in understanding how people effortlessly guide their eyes during reading (Rayner, 1998). The first computational model of eye movement control in reading was published by Morrison (1984), whose model was further refined by Rayner & Pollatsek (1989), Pollatsek & Rayner (1990), and finally Reichle et al. (1998), who nominated the model as E-Z Reader (see Reichle et al., 2012 for the latest version). The model computes fixation durations and saccade length as a response to sentence input, that is, a sequence of words. The model consists of algebraic formulas to simulate visual, oculomotor, and attentional lexical processes with both a priori-given and contextually determined word properties such as eccentricity (i.e., distance of a letter from a fixation point in visual angles), word length, frequency, and predictability.

The E-Z Reader made a few critical theoretical assumptions that stimulated much research: While all words and letters in a perceptual span are visually and inattentionally processed in parallel, attention is a prerequisite for lexical processing, and attention is shifted strictly serially between the words. Attentional lexical processing is divided into early L1 ("familiarity check") and late L2 stages (lexical recognition). The time required to complete the L1 stage depends mainly on the eccentricity of letters of the foveal word, its word frequency, and predictability. In contrast, the L2 stage depends only on the latter two. Completion of the L1 stage informs the system that the word will be recognized soon, and the preparation of the next saccade can be started. Completion of the L2 stage for a foveal word allows attention to shift to the next word – whenever the L1 of the next word is completed fast enough to cancel the saccade program, the next word is skipped, and saccade is targeted to the center of the subsequent word, n + 2. The alternative for this serial-attention model was provided by the SWIFT model (Engbert et al., 2002; Seelig et al., 2020), which assumes parallel attentional processing of multiple words within a perceptual span. More specifically, the number of words subject to attentional gradient is dynamically changed according to foveal processing demands, which, in turn, are determined by word properties such as frequency and predictability. Thus, several words are attended in parallel to compete for recognition. Saccade is targeted to the most activated word, which is not yet recognized (SWIFT; Seelig et al., 2020). Glenmore (Reilly & Radach, 2006) extended the dynamic field activation to the letter level, meaning that the letter perception also depends on the attention gradient.

The goal of the aforementioned eye movement control models was to estimate the word processing demands in a perceptual span and their influences on saccade targeting, not to model word recognition as an activity of mental representations. Only recently, the first attempts to integrate visual word recognition and models of eye movement control have arrived, namely the OB1 model for word-spaced orthographies (Snell et al., 2018) and (Li and Pollatsek, 2020); hereafter LP20 for Chinese reading. OB1 and LP20 models can be seen as extensions of the IAM to conduct successive word recognition, in which eye movements serve as a mechanism to control the amount of new textual information intake. In our view, the core theoretical issue that these models need to resolve is the transsaccadic integration (Higgins & Rayner, 2015; Pollatsek et al., 2015), that is, how visual words are preactivated during parafoveal preview, how these pre-activations facilitate later recognition of foveal words, and how both of these activations relate to saccade targeting during reading. In principle, the activations may occur in a single neural network simultaneously (OB1), sequentially in time (LP20), or occur as an information transmission between different spatial receptive fields (Kaiser et al., 2019; Golomb, 2019), e.g., activated representation for a word CAT at the parafoveal field is transferred into the foveal field during the presaccadic attention. We will next take a closer look at the state-of-the-art integrative and word recognition models.

2. State-of-the-art models

Integrative models. OB1 (Snell et al., 2018) was the first model that incorporated the IAM type of a word recognition module into an eye movement control model. Two major theoretical steps forward were made. Firstly, the model formally expressed how parallel word activations may be managed; that is, their activation may occur within a single IAM type of network and endure across a saccade to achieve transsaccadic integration. Although visuo-attentional constraints heavily favor the recognition of foveal words over the parafoveal ones, in principle, this design enables an easy parafoveal word to be recognized prior to a difficult foveal word leading to an error in the word order. For instance, a sentence such as "hen saw the fly" may be read as "hen the saw fly." However, the rate of word order errors was kept under control by another mechanism of theoretical significance, that is, the spatiotopic representations of words in a sentence, consisting essentially of a list of word length slots extracted from the low-level vision. Any word exceeding the recognition threshold is placed on a free slot matching for length, therefore preventing most word order errors from occurring. The overall design assuming some vulnerability for making word order errors is based on the findings that humans easily make word order confusions during reading (Snell & Grainger, 2019). Yet, others have argued that making such word order errors stems from the higher-level sentence parsing processes (Huang & Staub, 2022). Other advancements of the OB1 model include open bigram coding to achieve flexible coding of letter positions and to supplement visual acuity gradient with visual crowding factors to more realistically model the visual encoding of letters.

The OB1′s representational scheme of coding spatial order and length of words works for word-spaced orthographies. However, readers in non-spaced orthographies must first segment words from text as any character may start, belong, or end a word in the non-spaced orthography. Solving this segmentation challenge was the major theoretical advancement of the LP20 model (Li & Pollatsek, 2020) for reading Chinese. Similarly to OB1, the model incorporates a single IAM network with a predetermined lexicon. Visual acuity and attention gradient provide strong processing benefits for foveal characters, which initially activate word representations of various lengths and overlap with each other. The foveal word is recognized first, determining the next word's beginning. Thus, the activation of a foveal word precedes the activation of a parafoveal word within a single IAM network. Another potential theoretical advancement of this approach discussed by the authors is that the model may be extended to simulate the morphological decomposition universally, that is, also in word-spaced orthographies.

Word recognition. The IAM incorporated by OB1 and LP20 accounts for the orthographic processing, whereas the more comprehensive word recognition models (CDP++, DRC) also cover the mapping to phonology. A fundamental aspect of this mapping is the ability to phonologically decode novel words by using systematicity in the grapheme-phoneme relationships – a property that varies greatly across orthographies (Katz & Frost, 1992). The decoding ability provides a powerful self-teaching mechanism to acquire word representations, greatly facilitating word recognition (Share, 1995; 2008). The dual-route models (DRC, CDP++) assume that readers establish direct connections from the orthographic word representations to their phonological correspondents, while novel words are read by the serially working grapheme-phoneme conversion mechanism. Any word is processed in both routes, with the faster one dominating the output. The CDP++ first parses serially and syllable slots for stress assignment, then maps graphemes to phonemes in a serial fashion. Although less discussed, the routes may also have some interaction: Early in the processing, top-down feedback from IAM may fasten the encoding of novel letter strings resembling represented words in the orthographic lexicon.

The model can simulate a wide range of phenomena being arguably currently the most comprehensive model of word recognition.

Some authors have attempted to construct a single neural network model to recognize all types of words. The state-of-the-art of this approach is the recursive autoencoder model developed by Sibley et al. (2008, 2010, 2012). The model learns transitional probabilities between grapheme-phoneme pairs in trained words and serially produces a sequence of phonemes as a response to to word input. By doing so, more predictable phonemes receive stronger activation and can thus be considered to be read faster. The model can simulate a wide range of word recognition phenomena, including pseudoword reading, word frequency and length effects and their interaction, and several 'graphotactics' effects, including the number of syllable effects. Overall, the model can to the strong phonological view of reading (Frost, 1998), which assumes that word recognition always includes some sublexical phonological computation.

3. Some remaining issues for integrative modeling

Visual processing. All the existing models assume that letter encoding speed is dependent on eccentricity. Consequently, the models predict the word length effect to stem mainly from the visual processing level. This is a departure from word recognition models, which assume that the word length effect stems from the serial decoding. In fact, eccentricity has been found to influence strongly both the accuracy and speed of word and letter recognition (Staugaard et al., 2016; Veldre et al., 2023; Xiong et al., 2019), while the potential causality between accuracy and speed is unclear. It may be that a lower signal-to-noise ratio at higher eccentricities leads to longer consideration before the response. Thus, integrative models may consider modeling the effect of visual eccentricity on letter encoding merely as a confidence level rather than speed.

Transsaccadic integration. OB1 and LP20 assume parallel word activations to occur within a single IAM network and that only the recognized words are attributed to the spatiotopic sentence representation. This solution is limited in relation to the knowledge that the brain constantly encodes spatial information of objects in both egocentric and object-centered space and integrates this spatial information with object identity information (Committeri et al., 2004). In the egocentric coding, the activation of object representations takes place in a specific receptive field, e.g., a foveal word in one spatial receptive field and a parafoveal word in another field, while the transmission of information between these fields occurs via the presaccadic attention. The object-centered view can be seen as a working memory representation containing information about positions of each object in the environment (e.g., word locations within a sentence, paragraph, and page). Integrative modeling efforts should continue to pursue both ego- and object-centric spatial representational schemes for letters and words.

Attention allocation. Among the existing eye movement control models of reading, there are two fairly different concepts of attention. The serial attention E-Z Reader model assumes that a single word is lexically processed at a time. Instead, in gradient models (SWIFT, Glenmore), attention modulates the visual processing, thus affecting which letters and words are processed and at what rate. Based on the accumulated support for parallel lexical processing of multiple words (Snell & Grainger, 2019a, 2019b), and the other hand, the recent eye movement findings of orthographic activations to precede grapheme-phoneme decoding (Hautala et al., 2021), we believe there is a third credible concept of attention worth of investigation, namely that only the late, decoding stage of the word recognition requires attention. This possibility is also supported by the observed qualitative difference between the parafoveal preview and the foveal processing, where the previous studies have not found evidence for serial decoding during preview – the previewed word seems to be processed as a whole with both initial and final letters contributing roughly equally on lexical activation (Gagl et al., 2014; Briihl & Inhoff, 1995) and previous fixation duration being unaffected by parafoveal word length (e.g., Hautala et al., 2011; Hautala and Loberg, 2015) or other word properties (see Brothers et al., 2017).

Phonological decoding. Visual word recognition theories try to explain why word length effect is stronger for less frequent words (e.g., Weekes, 1997). In the CDP++ model, letters are first encoded in parallel, which, in turn, activates the word representations within the IAM (McClelland & Rumelhart, 1981) producing a word frequency effect. At the same time, the slower decoding route decodes the word according to the grapheme-phoneme correspondences, producing a length effect for words not included in the orthographic lexicon. A single-route alternative (Sibley et al., 2012) decodes the letters to phonemes serially, and therefore conflict with the parallel nature of vision (Nassi & Callaway, 2009). However, studying 3rd to 4th Grade students (9–10 years) across a continuum of reading fluency in a transparent Finnish orthography, Hautala et al. (2021) showed that the word frequency effect precedes the frequency and length interaction in readers' eye movement measures. Among fluent readers, a quantile regression analysis of the first fixation duration revealed a word frequency effect within five quantiles, and the interaction effect only appeared within 7.5 quantile. Among less fluent readers, the progression was delayed, with the first fixation duration showing only a word frequency effect and the refixation duration the interaction effect. Further, it was shown that the interaction effect is also present in the number of first-pass fixations in these readers (Hautala et al., 2023; see also, Huestegge et al., 2009). These results of word frequency effect preceding length effects are suggestive that activation of orthographic word representations may precedes and constitute an input to the decoding procedure. The authors labelled this as a dual-**stage** view of word recognition according to highly activated letter sequences may facilitate grapheme-phoneme conversion relative to weakly activated ones (Sibley et al., 2008; Sibley et al., 2010; Sibley and Kello, 2012). The design of a computational implementation for this view will be presented in this article.

Developmental dyslexia and reading fluency. Developmental dyslexia is a specific difficulty in developing typical reading skills, which stems from the interplay of genetic and environmental factors. It affects first language development, interfering first with reading acquisition and later with reading fluency development (Yang et al., 2022). Hautala et al. (2021) found that reading fluency was mostly explained (35 % of variance) by first fixation duration, while 14 % of the additional variance was explained by the larger word length and frequency effect in refixation durations, reflecting the decoding process. These results were interpreted to suggest a principal deficit in early visuo-orthographic processing (e.g., letter encoding) and secondary difficulties in decoding efficiency. In contrast, the early lexical processing seems to be intact – at least for those words included in the reader's orthographic lexicon. These results align with the findings derived from another transparent orthography, Italian (Zoccolotti et al., 2009). Thus, deficits in letter encoding and decoding should be considered when simulating reading fluency (Saksida et al., 2016; Ziegler et al., 2019). These deficits may partly result from less reading practice.

Discrete vs. dynamic control. Another major design choice of an integrative model is whether forward saccades are under discrete or dynamic control (Hautala et al., 2022). The discrete control view assumes a selection process to underlie which word a saccade is targeted (e.g. n, n + 1, or n + 2). Instead, the dynamic control means that the properties of the current and the next word adjust the saccade length in a continuous manner. The prevalent models (E-Z Reader, SWIFT, Glenmore, OB1) in word-spaced orthographies have all assumed discrete control. Accordingly, word recognition processes govern which word (its center) becomes the saccade target. However, the saccade length is further subject to bias towards a preferred saccade length (a systematic error) and random oculomotor error. In contrast, dynamic adjustment models have been developed for non-spaced Chinese orthography (Liu et al., 2018; Li & Pollatsek, 2020). According to these models, saccade length increases as a function of parafoveal preprocessing of the next word, which, in turn, is facilitated by the predictability and frequency of the next word.

However, Cutter and colleagues (2018a,b) revisited the targeting of interword saccades during reading in a word-spaced English orthography. They found that instead of relying on preferred saccade length, readers program forward saccade length based on center distance between the two adjacent words and, additionally, based on the actualized landing position on the first of these words. They also discussed that the center distance between the words n + 1 and n + 2 may already be perceived from the word n, resulting in a preliminary saccade plan, which is then corrected according to the actual landing position. This center-based saccade length account (CBSL) received further support from Hautala et al. (2022). They showed that by incorporating natural constraints (minimum and maximum) on the planned saccade length, the CBSL can be generalized to explain the word length effects in landing position, word skipping, and refixation probability. In addition, they simulated landing position distributions in high skipping probability conditions both for discrete and dynamic control models. As expected, the discrete model predicted bimodal and the dynamic model a unimodal landing position distribution over the words n + 1 and n + 2 in a high skipping probability condition. In favor of the dynamic control view, empirical data showed a unimodal landing position, meaning that most of the skipping saccades landed on the word space or the beginning of word n + 2. Finally, a recent study showed that a dynamic adjustment model was superior to a discrete control model in explaining saccade targeting in a non-linguistic Landolt C task, whose visuo-spatial requirements closely mimic reading (Xia et al., 2024). Thus, there is a need to develop dynamic adjustment models also for word-spaced orthographies. See Table 1 for a summary of the discussed conceptual differences between the OB1, LP20, and our proposed model.

4. A neural network approach to integrative modeling

In the previous chapter, we identified several remaining issues to be tackled by integrative models of word recognition and eye movement control. As many of the issues raised were fundamental, it is justified to consider also developing a new model from scratch. For such a task, a generic neural network modeling has many favorable qualities (McClelland, 2010). By generic, we mean the attempt to directly translate neural network activation into behavior, with minimal a priori constraints. In this translation, the identity of the most activated node (potentially exceeding a predefined recognition threshold) can be used as an index of accuracy, and its activation values as an index of processing speed (Dandurand, Hannagan, & Grainger, 2013; Sibley and Kello, 2012).

Activations emerge when a model with a priori-defined architecture is trained with suitable stimulus materials such as a list of words. For example, the word frequency effect emerges when some words are presented to the network more often than other words, strengthening their corresponding representations. Note that different word frequencies can also be modeled by using weighted loss functions (Kärkkäinen, 2002). As another example, short-term learning can be seen as the formation of new representations, whereas the long-term development can be seen as a change in the global activation dynamics of different layers (Shultz, 2017). Moreover, reading disorders like developmental dyslexia may be simulated as an added noise to specific layers (Perry et al., 2019) or as a lesser amount of training. These examples illustrate that the generic neural network approach has a promise to lead to a highly parsimonious yet wide-scope model of reading. We will next present our conceptual model based on this approach.

Table 1
Conceptual summary of the existing and the here proposed integrative models.

Concept	OB1	LP20	Dual-stage NN
Visual processing	Time-costly	Time-costly	Time-independent
Transsaccadic integration	Spatiotopic	Ordered	Receptive fields
Attention allocation	Parallel	Parallel	Parallel-to-serial
Word recognition	Orthographic	Orthographic	Orthographic & Phonological
Saccade control	Discrete	Discrete	Dynamic
Learning and development	Model tuning	Model tuning	Trainable

5. The proposed model design

Like most previous models (E-Z Reader, SWIFT, Glenmore, OB1, LP20), we assume that word recognition is the driving force of eye movements. Therefore, the model's core will be a foveal word recognition module (Fig. 1) consisting of neural network layers specializing in different subprocesses laid down by the dual-stage view of word recognition (Hautala et al., 2021). As the next step, this module will be extended to process several words simultaneously, which requires modeling the system's visual front-end (acuity, crowding), attentional mechanism, and specifying how multiple word activations are managed. The third step is to specify how activations at different network layers contribute to progressive saccade length as specified by the dynamic adjustment mechanism of saccade lengths (Hautala et al., 2022). Only after establishing these functionalities can we start implementing detailed saccade planning procedures following the formulas provided by the previous models. Thus, currently, the scope of the design is limited to model foveal and parafoveal word activations and the amount of (saccadic) shift in the input text.

6. Implementation plan

Dual-stage word recognition. The dual-stage module of word recognition is planned to have a 'hybrid' autoencoder architecture (Attia et al., 2017: Khamparia et al., 2020), in which parallel encoding layers are followed by serial recurrent (i.e., Long Short Term Memory, LSTM) decoding layers.

For the encoder part, we are planning to follow (Dandurand et al., 2013) , who trained a network with words presented in variable locations – as it happens during continuous reading also. Their model consists of retinotopic input layer coding letters and their positions. The inputs are then mapped into another layer via hidden layer, resulting in a word-centered letter string representation (letters in their positions). These representations are then mapped on the lexical representations at the output layer. More specifically, the input layer consisted of location-specific detectors for each letter, with each location subject to visibility scaling. Visibility scaling was used to reduce the excitability of letter detectors that receive activation from farther eccentricities and word-internal letter positions, thus simulating acuity and crowding effects. The word-centered representation codes letter in their positions in a word (e.g., #C###, ##A##, ###T#; Dandurand et al., 2013; Sibley et al., 2012). In this scenario, the inverse sum of the most activated nodes at each position might be taken as a proxy of word activation time.

For the decoding part, we will follow Sibley et al. (2012) by incorporating a recurrent layer coding highly context-specific transitional probabilities among the adjacent letter-phoneme conjunctions (e.g., between #C###, ##A##, ###T#). Again, the inverse sum of the most activated phoneme at each step might be taken as a proxy of decoding time for a particular word (Sibley et al., 2012). The sum of word activation and decoding would then constitute a proxy for word recognition time.

In a trained model, the output activations would depend both on the model's exposure to the input and resembling letter strings (Fig. 2). If the input (e.g., CAPITAL) is a frequent word and thus well learned, all letters in the word-centered representation are strongly activated. If the transitional probabilities between letters are high, phonemes will also be highly activated, and decoding will be deemed fast. If there is some anomaly in the input (e.g., CAPIATL), word-centered activation may be reduced only to some extent, whereas decoding of the transposed letters

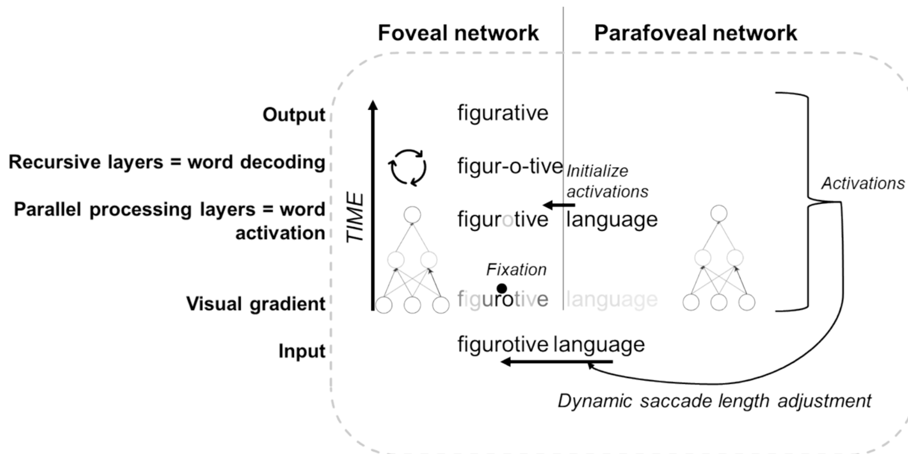

Fig. 1. Schematic diagram of the proposed model design. The input is represented at the visual gradient favoring the foveal word. Letter shading represents the activation levels and hyphens the decoding difficulty – there is a deliberate typo in the foveal word to illustrate the activation dynamics. Two successive letter strings separated by the word space activate word-centered letter string representations independently at foveal and parafoveal networks (Reichle & Schotter, 2020), but only the foveal word is entered into the decoding layer. Saccade length is determined by activations in all layers with the visual gradient setting minimum and maximum saccade lengths. Within these limits, saccade is targeted towards the center of the parafoveal word, while lexical activations contribute to fine-tuning of saccade length. At the beginning of fixation to a new word, the foveal orthographic network is initialized by the previous parafoveal network activation.

Recursive:	C-A-P-I-T-A-L	C-A-P-I--A--T--L	A--T--P--C--A--I--L
Word-centered:	##CAPITAL##	##CAPIATL##	##ATPCAIL##
Retinotopic:	####CAPITAL##	####CAPIATL##	####ATPCAIL##

Fig. 2. Schematic diagram of representational levels. Letters of visual words are first encoded retinotopically, which input activates object-centered (i.e. word) representation to be further decoded into phoneme sequences. Font darkness represents activation levels and number of hyphens decoding difficulty level. Hash signs represent empty spaces. In this example, the fixation location would be at the third letter.

may be affected more. When the input is genuinely novel (e.g., ATPCAIL), all activations would be low both in word-centered representation and during decoding.

In simulation studies, the emphasis will be study the transition from parallel encoding layer to the recurrent network. For this purpose, simulations with relatively short words suffice (e.g., 2 to 8 letters). The model performance will be evaluated by studying its ability to correctly recognize words while producing the "dual-stage" pattern of processing sequence, that is, the word frequency effect followed by the word frequency and length interaction effect (Hautala et al., 2021). The model's ability to produce other key benchmark effects, such as transposed letter effects, must also be demonstrated (Dandurand et al., 2013).

Integrative model. In the next phase, the dual-stage model will be extended into an integrative model. Central to this design is that the retinotopic input layer feeds two independent word-centered networks corresponding to foveal and parafoveal words (see Reichle & Schotter, 2020), with only the foveal network feeding into the subsequent decoding layer. Network activation update mechanisms determine how word activations are integrated and coordinated across saccades. Initially, we plan to implement these as rules based on realized landing position, although we acknowledge the need to develop a seamless neural network solution. The sum of word and decoding activations would be taken as the main proxy of fixation duration, now supplemented with network update and decoder initialization costs and preview benefit. However, in later improvements of the model, we will consider modeling temporal activation dynamics, that is, how fast the activation rises within the network, as is done in previous models (SWIFT, Glenmore, E-Z Reader). Note that simulating word predictability effects is yet out of the scope of the current design (see Limitations –section in the Discussion). Next, we will describe the planned functioning of the word recognition network and the determination of saccade length in the integrated model.

The input layer with the visibility scaling is extended to cover the whole perceptual span (i.e., 18 locations) to determine to which extent each letter in the perceptual span is excitable. This produces a strong activation benefit for foveated over parafoveal word in the subsequent hidden layers. For simplicity, word spaces determine to which network a letter string is mapped. During an initial fixation into a word, the decoding layer produces an output letter string only for the encoded part of the letter string (i.e., part of letter string exceeding some a priori defined threshold; Dandurand et al., 2013), which is mainly constrained by the visibility functions. Thus, when the word is short, the decoding will be completed during a single fixation, but when the word is long, only its beginning part is being decoded during the first fixation. This property limits the influence of word length on first fixation duration and the other hand, contributes to inducing a strong word length effect on refixation probability (Hautala et al., 2021).

To complete the decoding of a long word, we need to update network activations. When a refixation occurs, the activations in the foveal and parafoveal networks should be *maintained and refined* at the new fixation position. Because decoding has already begun and foveal and parafoveal words remain in their respective networks, no network update and decoder initialization costs may be executed for refixations. In addition, the word-end letters likely gained some activation during the first fixation, producing a preview benefit. These factors are expected to cause refixations to be generally around 20–30 ms shorter than the first fixations (Hautala et al., 2023; Loberg et al., 2019) and to be relatively more affected by the decoding process (i.e., frequency x length interaction) than the first fixation (Hautala et al., 2021; 2023). Because letter activations of a word are set only to increase, not to decrease, letters encoded or decoded during an earlier fixation are remembered, even if they are not excited anymore at the current fixation location.

In the case of interword saccade leading to first fixation, the foveal network activations are planned to be first *replaced (or initialized)* with parafoveal network activations, resulting in a preview benefit but also a network update cost. The activations would then be refined during the fixation and fed into the decoder network with the initialization cost.

In the case of a skipping saccade, the network activations are planned to be first updated as if the fixation would have landed on a skipped word. As soon as the decoding of a skipped word is completed, another network update is conducted without a saccade, allowing thus foveal processing to catch up with the current eye location. This additional network update should produce a constant skipping cost on fixation duration after skipping (Hautala et al., 2022; Reichle & Drieghe, 2013).

We will consider several ways to implement the dynamic adjustment mechanism of the saccade lengths (Hautala et al., 2022). For fluent readers, a priori-defined saccade targeting mechanism would consist of calculating first the center distance between foveal and parafoveal words and adjusting this according to the launch distance towards the parafoveal word. Accordingly, when the launch distance is large, the saccade length would be shortened, leading more likely to a refixation. On the other hand, when the launch distance is small, the saccade length would also be shortened to secure an additional visual sample of the beginning of the next word or a word that will likely be skipped. As a more generic and arguably more realistic scenario, we will also pursue training the model to minimize the word recognition errors and number of fixations by adjusting the saccade length. Given the visibility constraints, this should lead to the emergence of the above-described eye movement behavior, that is, targeting word centers within the limits of minimum and maximum saccade lengths, securing refixations to long words, and securing visual sampling of word beginnings or overshoot, i. e., skipped words.

Model specification. Following Dandurand et al. (2013), the input layer would consist of binomial detectors for all letters at each position, and the subsequent word-centered hidden node activations would be calculated as the sum of input [0, 1] × learned weight [0–1] × visibility constraint [0–1] factor. We apply the same visibility constraint function used in the OB1 model (Snell et al., 2018).

Fixation duration will be determined by the following type of formula,

$$FD = PA_{F-1,W} + \Delta WA_{F,W} + \Delta WD_{F,W} + U_{F,W} + D_{F,W} \tag{1}$$

where the subscript F refers to the current fixation and W identity of a word in a sentence. Here, PA_{F-1} is word-centered letter activations of a currently fixated word in the parafoveal network during previous fixation, $\Delta WA_{F,\ W}$ is the increase in word-centered letter activations of currently fixated word in foveal network during current fixation, $\Delta WD_{F,\ W}$ is the change of activation in the recurrent network output layer for the currently fixated word, the constant U is an update cost, and D is the initialization cost of the decoder for first fixation. The magnitude for each cost is expected to be ∼ 10 ms.

Following Sibley et al. (2010), activation of the word-centered representation is

$$\Delta WA = s_a \times \left[\sum (1 - \Delta a_p)\right] \tag{2}$$

where scaling factor s_a transforms activation values on the observed scale of fixation durations (∼150 ms before adding other terms in the formula 1), a_p is the activation of an activated output unit for a letter position, and summation is across positions. In the trained model, the difference between novel and highly trained words should be < 25 ms (Hautala et al., 2021).

Similarly, activation of the decoding network is computed as

$$\Delta WD = s_d \times \left[\sum (1 - \Delta a_r)\right] \tag{3}$$

where s_d is a scaling factor, and a_r is the output unit's activation on a sequence step r, and the summation is across steps. The r is determined by the number of input letters exceeding a threshold for being encoded. The magnitude of ΔWD should be < 10 ms per letter, depending on how well the model is trained to read a particular word (Hautala et al., 2011; 2021).

The parafoveal activation is computed also similarly,

$$PA = s_b \times [\left(\sum (1 - \Delta a_b)\right] \tag{4}$$

where the scaling factor s_b transforms the values for the observed scale of known preview benefit (∼20–30 ms; Rayner et al., 2010), a_b is the activation of an activated output unit for a letter position, and summation is across all positions.

In addition to the above-defined raw activation measures, we will also consider confidence measures (Sibley et al., 2012), which quantify the winning node's marginal to its competing nodes.

Initially, the saccade length (SL) will be determined by the formula provided by Hautala et al. (2022):

$$SL = 3.85 + 0.54 \times CD - 0.019 \times CD^2 + 0.32 \times LD - 0.032 \times LD^2 - FD \times S1 - PA \times S2 \tag{5}$$

where CD is the distance between centers of foveal and parafoveal words, LD is the launch distance to the beginning of a parafoveal word, and FD and PA represent foveal and parafoveal processing processing difficulties, both scaled to correspond to the small observed lexical influences on saccadic length (Albrengues et al., 2019; Radach et al., 2008; Radach et al., 2004; White et al., 2018).

Modeling technique. After successful initial testing with a highly restricted set of items (Dandurand et al., 2010), the model will first be trained with a representative sample of Finnish words (Huovilainen, 2018) before simulations with different languages. To enable a direct comparison with the preceding work, the sentence corpuses used to train the previous models will be considered (e.g., Schilling et al., 1998). For the integrative model, fixation-specific activation values and landing positions provide a basis for calculating word-specific accuracy and speed indexes for different eye movement metrics, such as skipping probability, first fixation duration, refixation probability, refixation and gaze duration. We will first focus on simulating word frequency and length effects and validating them against observed effects and simulation results of previous models.

Technically, we plan to build all models and networks in Python 3.10.4 using TensorFlow version 2.11., or newer. Following Dandurand et al. (2010), we plan to initialize the weights of our network randomly and then use stochastic gradient descent with the Adam optimizer (Kingma & Ba, 2015) to find the optimal weights. Stochastic gradient descent calculates the error gradient for the current state of the model using points from the training set and updates the weights of the network through backpropagation. One of the most important hyperparameters hereby is the learning rate (i.e., the quantity of how much the weights are updated in each step). The previous work (Dandurand et al., 2010; Dandurand et al., 2013) used a learning rate 0.1. We will also experiment with smaller learning rates up to 0.0001 (such as Saarela & Georgieva, 2022) to find the globally optimal set of weights. See https://github.com/kirilkhalil/reread_neural for the current status of the modeling work.

7. General discussion

The proposed model here has many similarities, differences, and limitations in relation to previous models.

Similar to many previous models (E-Z Reader, SWIFT, Glenmore, OB1), the proposed model design assumes that the driving force of saccade length computation is the goal of optimizing word recognition. Concerning the fundamental debate of attentional processing, i.e., whether attention is allocated to words serially (E-Z Reader) or in parallel (SWIFT, Glenmore, OB1, LP20), the proposed model would be a hybrid one, that is, parallel-to-serial model. In the terminology of E-Z Reader, we propose that only L2 lexical processing requires attention, and in the current design, it consists only of grapheme-phoneme conversion. However, whereas the E-Z Reader assumes seriality for both L1 and L2 stages, our model assumes (similarly to SWIFT, Glenmore, OB1, LP20) that the early lexical processing L1 may occur in parallel for several words. Yet our model differs also from the parallel models in two important ways: First, the parallel processing stage maintains the spatial order of the words by separating the processing of foveal and parafoveal words on two independent but interlinked neural networks. This is different from the OB1 and LP20, in which parallel word activations occur in a single neural network, thus allowing currently fixated and upcoming words to compete for recognition.

Moreover, while the OB1 and LP20 assume that word recognition is a single process, we assume a two-stage process in which orthographic processing precedes phonological decoding of a foveal word. This phonological decoding is not included in any of the previous models among word length effects originate mainly from the longer time required to encode letters at farther eccentricities. However, the seriality is gradual, meaning it can be very fast or slow, depending on how well the model is trained to recognize a particular word. Therefore, in practice, it may turn out that the length effect among short high frequent words may be minimal and, therefore, functionally correspond to direct lexical associations from orthography to phonology. Nevertheless, the model conforms to the strong phonological view of reading (Frost, 1998), stating that word recognition always involves some sublexical phonological computation.

Perhaps the most fundamental difference between the present model design and existing models concerns the nature of visual processing. While the previous models assume that letter encoding time increases as a function of eccentricity, the present design assumes a decrease in activation levels. More empirical research is needed to unravel the causal relationship between the influence of viewing eccentricity on letter encoding accuracy and speed (Staugaard et al., 2016; Xiong et al., 2019; Veldre et al., 2023). In the present model design, activation levels provide a basis for calculating processing speed indexes, which could produce an early word length effect, yet complex mapping from input letter detectors to word-centered representations may suppress such effects. In addition, unlike in previous models, no extra time would help encode eccentric letters.

Theoretical strength of the design. The proposed model is designed to address several central conceptual challenges of integrative word recognition and eye movement control models.

Concerning word recognition, the proposed model attempts to implement the recently established dual-stage view of word recognition (Hautala et al., 2021). The first of the two stages consists of orthographic processing relying on widely accepted parallel distributed processing of letters to activate orthographic word representations (McClelland & Rumelhart, 1981), while the second phonological processing stage consists of serial production of phonemes (Sibley et al., 2012). With this architecture, the model should be able to reproduce the pattern of results signature for dual-stage processing, that is, the word frequency effect followed by the word frequency and length interaction in fixation durations (Hautala et al., 2021). The dual-stage model provides a credible alternative to avoid the pitfalls of dual-route models ((Coltheart et al., 2001; Perry et al., 2010): The dual-route view was not grounded on time-course information about word recognition, and consequently, it proposed two relatively independent reading routes. However, no evidence has been established that individuals would show bimodality in their reading times of words, with one distributional peak reflecting the working of the direct route and the second peak reflecting the working of the phonological route.

Among the integrative models, the main theoretical contributions of the proposed design are to provide new hypotheses both for the attention allocation between words and transsaccadic integration processes. The fully serial attention allocation model, E-Z Reader, as introduced, equalizes lexical with attentional processing. This produces strong time constraints for the system as parafoveal preprocessing can only begin when lexical processing of the foveal word is completed, but saccade is not yet launched. This contrasts with accumulated evidence of parallel lexical level of processing of two or more words (Snell & Grainger, 2019). On the other hand, parallel attention models (SWIFT, Glenmore, OB1) set successive words to compete with each other to be recognized, therefore allowing for making word order errors. Further, it is unclear how the dynamic attention field should be implemented. The parallel models suggest that effective vision enlarges when text gets easier. Another realistic option is that text difficulty mainly increases the foveal load on lexical processing, whereas the visual field and possibly also parafoveal preprocessing would remain largely unaffected by the text difficulty (Brothers et al., 2017; Vasilev & Angele, 2017). The proposed design equaling phonological decoding with attention aligns with this conceptualization: When a word is difficult, decoding takes more time. However, such prolonged fixations do not translate into increased preview benefit, because parafoveal activations are calculated once for each fixation. These considerations suggest the need to study the proposed parallel-to-serial attention allocation mechanism.

A major theoretical issue in the development of integrative models is the management of parallel word activations. Both existing solutions (OB1, LP20) have employed a single IA network in which foveal and parafoveal words are activated and possibly compete with each other for recognition. This induces a risk of making word order errors and neglects the availability of spatial information, according to how the brain organizes perception (Committeri et al., 2004; Kaiser et al. 2019; Golomb, 2019). Our suggested solution is to maintain the receptive fields of the current and next word and allow separate lexical competition within both fields. This does not allow for making word order errors because only the foveal word is decoded, and parafoveal activation, no matter how strong, is only a prediction, not full recognition. This view is not only compatible with the brain knowledge of spatial field-specific processing (Committeri et al., 2004; Kaiser et al. 2019; Golomb, 2019) but also matches with the generally accepted view of cognition as predictive coding according to predictions can greatly facilitate recognition but still needs to be confirmed with sensory information of sufficient quality (Friston, 2018).

Limitations. The proposed design includes several limitations in relation to both the previous models and the foreseen desirable qualities. First, the design still lacks saccade planning processes other than the computation of progressive saccade length. In other words, the here proposed model will not be able to simulate the time course of saccade planning, including its labile and non-labile stages.

Relatedly, perhaps the major theoretical limitation of the current design relates to presaccadic attention. As reviewed, neurological evidence indicates that the foveal receptive field of the visual cortex becomes sensitive to parafoveal features prior to saccade launch (Kaiser et al. 2019; Golomb, 2019). In the current design such anticipatory network activation update processes are not covered, as we believe these finesses should be refined only after the current design is confirmed to be functional.

Another limitation of the current design is the lack of a mechanism for correcting large deviations from the optimal landing position with a rapid progressive or regressive saccade aimed toward the word center (O'Regan and Jacobs, 1992). Neither does the proposed model simulate regressive saccades resulting, e.g., from difficulties in integrating the current word with the preceding sentence context (E-Z Reader) or having not recognized the preceding word (SWIFT, Glenmore). For the present, we leave these aspects unmodeled because they are mostly related to error components of the models, and we will first need to establish which error components our model needs in the first place. For example, the oculomotor error component is traditionally assumed to be a huge one, and in the dynamic adjustment framework, word properties may partly explain this variability.

The current design covers transsaccadic integration but is only an initial step in modeling the working memory processes involved in reading. First, the current design does not include any object-centered spatial representation of the sentence. A step toward this direction has been made by OB1 with its spatiotopic representation. However, currently its role is merely to act as a helper for a word recognition module as it does not involve linguistic representation of word sequences guiding, e.g., regressive saccades to particular words in a sentence. Modeling such regression behavior would require extending the scope of the models to morphological, syntactic, and semantic processing, which already are within the scope of neural network models of natural language processing (Schomacker & Tropmann-Frick, 2021).

Finally, we did not lay down a detailed plan to simulate word predictability effects. This is because the currently presented design is a prerequisite and foundation for modeling such higher-order effects. Theoretically, such effects could be modeled by setting another recurrent network to predict a parafoveal word based on a foveal word (and its antecedents). The parafoveal lexical activation could then provide top-down excitation to parafoveal word-centered representations and thus contribute to preview benefit. In principle, this mechanism may be sufficient to explain the reading of predictable word chains, although comprehensive modeling of predictability would probably require relying also on language models (Schomacker & Tropmann-Frick, 2021).

Conclusions. We have laid down a detailed plan to construct an integrative model of eye movement control and word recognition in reading, relying on a generic neural network approach with a minimal number of parameters. This design also enables us to simulate learning and development in the future. Theoretically, the design proposes that foveal and parafoveal words are processed in parallel at the orthographic processing level, whereas only the foveal word is subject to phonological decoding requiring attention. Saccade length is adjusted dynamically to optimize word recognition, which is expected to lead to targeting word centers within minimum and maximum saccade lengths. We believe the proposed design provides a valuable addition to theoretical understanding and computational modeling of reading. We also hope that this article simulates new research in this area.

Declaration of generative AI and AI-assisted technologies in the writing process

During the preparation of this work the authors used the Grammarly Premium −tool (not the generative GrammarlyGO) in order to improve the language and readability of the article. After using this tool/service, the authors reviewed and edited the content as needed and takes full responsibility for the content of the publication.

CRediT authorship contribution statement

Jarkko Hautala: Conceptualization, Funding acquisition, Investigation, Methodology, Project administration, Resources, Supervision, Visualization, Writing – original draft, Writing – review & editing. **Mirka Saarela:** Methodology, Supervision, Writing – review & editing. **Otto Loberg:** Conceptualization, Writing – review & editing. **Tommi Kärkkäinen:** Conceptualization, Methodology, Supervision, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

No data was used for the research described in the article.

Acknowledgements

This research was funded by the Research Council of Finland (formerly Academy of Finland) with grants 317030, 319911 and 345639 to Jarkko Hautala. The authors report no conflict of interest. The authors thank the reviewers for their valuable comments.

References

Albrengues, C., Lavigne, F., Aguilar, C., Castet, E., & Vitu, F. (2019). Linguistic processes do not beat visuo-motor constraints, but they modulate where the eyes move regardless of word boundaries: Evidence against top-down word-based eye-movement control during reading. *PLoS ONE, 14*. https://doi.org/10.1371/journal.pone.0219666

Attia, M., Hossny, M., Nahavandi, S., & Asadi, H. (2017, October). Surgical tool segmentation using a hybrid deep CNN-RNN auto encoder-decoder. In 2017 IEEE International conference on systems, man, and cybernetics (SMC) (pp. 3373-3378). IEEE. doi: 10.1109/SMC.2017.8123151.

Briihl, D., & Inhoff, A. W. (1995). Integrating information across fixations during reading: The use of orthographic bodies and of exterior letters. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*(1), 55. https://doi.org/10.1037/0278-7393.21.1.55

Brothers, T., Hoversten, L. J., & Traxler, M. J. (2017). Looking back on reading ahead: No evidence for lexical parafoveal-on-foveal effects. *Journal of Memory and Language, 96*, 9–22. https://doi.org/10.1016/j.jml.2017.04.001

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. https://doi.org/10.1037/0033-295X.108.1.204

Committeri, G., Galati, G., Paradis, A. L., Pizzamiglio, L., Berthoz, A., & LeBihan, D. (2004). Reference frames for spatial cognition: Different brain areas are involved in viewer-, object-, and landmark-centered judgments about object location. *Journal of cognitive neuroscience, 16*(9), 1517–1535. https://doi.org/10.1162/0898929042568550

Cutter, M. G., Drieghe, D., & Liversedge, S. P. (2018a). Reading sentences of uniform word length–II: Very rapid adaptation of the preferred saccade length. *Psychonomic bulletin & review, 25*(4), 1435–1440. https://doi.org/10.1037/xhp0000416

Cutter, M. G., Drieghe, D., & Liversedge, S. P. (2018b). Reading sentences of uniform word length–II: Very rapid adaptation of the preferred saccade length. *Psychonomic bulletin & review, 25*, 1435–1440. https://doi.org/10.3758/s13423-018-1473-2

Dandurand, F., Grainger, J., & Dufau, S. (2010). Learning location-invariant orthographic representations for printed words. *Connection Science, 22*(1), 25–42. https://doi.org/10.1080/09540090903085768

Dandurand, F., Hannagan, T., & Grainger, J. (2013). Computational models of location-invariant orthographic processing. *Connection Science, 25*(1), 1–26. https://doi.org/10.1080/09540091.2013.801934

Davis, C. J. (1999). The self-organising lexical acquisition and recognition (SOLAR) model of visual word recognition (Doctoral dissertation, UNSW Sydney).

Davis, C. J. (2010). The spatial coding model of visual word identification. *Psychological Review, 117*(3), 713. https://doi.org/10.1037/a0019738

Engbert, R., Longtin, A., & Kliegl, R. (2002). A dynamical model of saccade generation in reading based on spatially distributed lexical processing. *Vision Research, 42*(5), Article 621636. https://doi.org/10.1016/S0042-6989(01)00301-7

Friston, K. (2018). Does predictive coding have a future? *Nature neuroscience, 21*(8), 1019–1021. https://doi.org/10.1038/s41593-018-0200-7

Frost, R. (1998). Toward a strong phonological theory of visual word recognition: True issues and false trails. *Psychological Bulletin, 123*(1), 71. https://doi.org/10.1037/0033-2909.123.1.71

Gagl, B., Hawelka, S., Richlan, F., Schuster, S., & Hutzler, F. (2014). Parafoveal preprocessing in reading revisited: Evidence from a novel preview manipulation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 40*(2), 588. https://doi.org/10.1037/a0034408

Golomb, J. D. (2019). Remapping locations and features across saccades: A dual-spotlight theory of attentional updating. *Current Opinion in Psychology, 29*, 211–218. https://doi.org/10.1016/j.copsyc.2019.03.018

Grainger, J. (2003). Moving eyes and reading words: How can a computational model combine the two?. In *The Mind's Eye* (pp. 457-470). North-Holland.

Hautala, J., Hyönä, J., & Aro, M. (2011). Dissociating spatial and letter-based word length effects observed in readers' eye movement patterns. *Vision Research, 51*(15), 1719–1727.

Hautala, J., & Loberg, O. (2015). Breaking down the word length effect on readers' eye movements. *Language, Cognition and Neuroscience, 30*(8), 993–1007. https://doi.org/10.1016/j.visres.2011.05.015

Hautala, J., Hawelka, S., & Aro, M. (2021). Dual-stage and dual-deficit? Word recognition processes during text reading across the reading fluency continuum. *Reading and Writing, 35*(3), 663–686. https://doi.org/10.1007/s11145-021-10201-1

Hautala, J., Hawelka, S., Loberg, O., & Leppänen, P. H. (2022). A dynamic adjustment model of saccade lengths in reading for word-spaced orthographies: Evidence from simulations and invisible boundary experiments. *Journal of Cognitive Psychology, 34*(4), 435–453. https://doi.org/10.1080/20445911.2021.2011895

Hautala, J., Hawelka, S., & Ronimus, M. (2023). An eye movement study on the mechanisms of reading fluency development. *Cognitive Development*. https://doi.org/10.1016/j.cogdev.2023.101395

Higgins, E., & Rayner, K. (2015). Transsaccadic processing: Stability, integration, and the potential role of remapping. *Attention, Perception, & Psychophysics, 77*, 3–27. https://doi.org/10.3758/s13414-014-0751-y

Huang, K. J., & Staub, A. (2022). The transposed-word effect does not require parallel word processing: Failure to notice transpositions with serial presentation of words. *Psychonomic Bulletin & Review, 1–8*. https://doi.org/10.3758/s13423-022-02150-9

Huestegge, L., Radach, R., Corbic, D., & Huestegge, S. M. (2009). Oculomotor and linguistic determinants of reading development: A longitudinal study. *Vision Research, 49*(24), 2948–2959. https://doi.org/10.1016/j.visres.2009.09.012

Huovilainen, T. M. (2018). Psycholinguistic descriptives. The Language Bank of Finland. Retrieved from: http://urn.fi/urn:nbn:fi:lb-2018081601.

Kaiser, D., Quek, G. L., Cichy, R. M., & Peelen, M. V. (2019). Object vision in a structured world. *Trends in cognitive sciences, 23*(8), 672–685. https://doi.org/10.1016/j.tics.2019.04.013

Katz, L., & Frost, R. (1992). The reading process is different for different orthographies: The orthographic depth hypothesis. In *Advances in psychology* (Vol. 94, pp. 67-84). North-Holland.

Khamparia, A., Pandey, B., Tiwari, S., Gupta, D., Khanna, A., & Rodrigues, J. J. (2020). An integrated hybrid CNN–RNN model for visual description and generation of captions. *Circuits, Systems, and Signal Processing, 39*(2), 776–788. https://doi.org/10.1007/s00034-019-01306-8

Kingma, D. P., & Ba, J. (2015). Adam: A method for stochastic optimization, in: Proceedings of International Conference on Learning Representations, 860. doi: 10.48550/arXiv.1412.6980.

Kärkkäinen, T. (2002). MLP in layer-wise form with applications to weight decay. *Neural Computation, 14*(6), 1451–1480. https://doi.org/10.1162/089976602753713016

Li, X., & Pollatsek, A. (2020). An integrated model of word processing and eye-movement control during Chinese reading. *Psychological Review, 127*(6), 1139. https://doi.org/10.1037/rev0000248

Liu, Y., Guo, S., Yu, L., & Reichle, E. D. (2018). Word predictability affects saccade length in Chinese reading: An evaluation of the dynamic-adjustment model. *Psychonomic Bulletin & Review, 25*, 1891–1899. https://doi.org/10.3758/s13423-017-1357-x

Loberg, O., Hautala, J., Hämäläinen, J. A., & Leppänen, P. H. (2019). Influence of reading skill and word length on fixation-related brain activity in school-aged children during natural reading. *Vision Research, 165*, 109–122. https://doi.org/10.1016/j.visres.2019.07.008

McClelland, J. L. (2010). Emergence in cognitive science. *Topics in Cognitive Science, 2*(4), 751–770. https://doi.org/10.1111/j.1756-8765.2010.01116.x

McClelland, J. L., & Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception: I. An account of basic findings. *Psychological Review, 88*(5), 375. https://doi.org/10.1037/0033-295X.88.5.375

Morrison, R. E. (1984). Manipulation of stimulus onset delay in reading: Evidence for parallel programming of saccades. *Journal of Experimental Psychology: Human Perception and Performance, 10*, 667–682. https://doi.org/10.1037/0096-1523.10.5.667

Nassi, J. J., & Callaway, E. M. (2009). Parallel processing strategies of the primate visual system. *Nature Reviews Neuroscience, 10*(5), 360–372. https://doi.org/10.1038/nrn2619

O'Regan, J. K., & Jacobs, A. M. (1992). Optimal viewing position effect in word recognition: A challenge to current theory. *Journal of Experimental Psychology: Human Perception and Performance, 18*(1), 185. https://doi.org/10.1037/0096-1523.18.1.185

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. https://doi.org/10.1037/0033-295X.114.2.273

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. https://doi.org/10.1016/j.cogpsych.2010.04.001

Perry, C., Zorzi, M., & Ziegler, J. C. (2019). Understanding dyslexia through personalized large-scale computational models. *Psychological science, 30*(3), 386–395. https://doi.org/10.1177/0956797618823540

Pollatsek, A., & Rayner, K. (1990). Eye movements and lexical access in reading. In D. A. Balota, G. B. Flores d`Arcais, & K. Rayner (Eds.), *Comprehension processes in reading* (pp. 143-163). Hillsdale, NJ: Erlbaum.

Pollatsek, A., Treiman, R., Cutter, M. G., Drieghe, D., & Liversedge, S. P. (2015). How is information integrated across fixations in reading? *The Oxford Handbook of Reading.* https://doi.org/10.1093/oxfordhb/9780199324576.013.11

Radach, R., Huestegge, L., & Reilly, R. (2008). The role of global top-down factors in local eye-movement control in reading. *Psychological Research, 72*(6), 675–688. https://doi.org/10.1007/s00426-008-0173-3

Radach, R., Inhoff, A., & Heller, D. (2004). Orthographic regularity gradually modulates saccade amplitudes in reading. *European Journal of Cognitive Psychology, 16*(1–2), 27–51. https://doi.org/10.1080/09541440340000222

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological bulletin, 124*(3), 372. https://doi.org/10.1037/0033-2909.124.3.372

Rayner, K., & Pollatsek, A. (1989). *The psychology of reading.* Prentice Hall.

Rayner, K., Castelhano, M. S., & Yang, J. (2010). Preview benefit during eye fixations in reading for older and younger readers. *Psychology and Aging, 25*(3), 714. https://doi.org/10.1037/a0019199

Reichle, E. D., & Drieghe, D. (2013). Using EZ reader to examine word skipping during reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 39*(4), 1311. https://doi.org/10.1037/a0030910

Reichle, E. D. (2021). *Computational models of reading: A handbook*. Oxford University Press.

Reichle, E. D., Pollatsek, A., & Rayner, K. (2012). Using EZ Reader to simulate eye movements in nonreading tasks: A unified framework for understanding the eye–mind link. *Psychological Review, 119*(1), 155. https://doi.org/10.1037/a0026473

Reichle, E. D., Pollatsek, A., Fisher, D. L., & Rayner, K. (1998). Toward a model of eye movement control in reading. *Psychological Review, 105*(1), 125. https://doi.org/10.1037/0033-295X.105.1.125

Reichle, E. D., & Schotter, E. (2020). A Computational Analysis of the Constraints on Parallel Word Identification. In *CogSci*. Retrieved from http://faculty.cas.usf.edu/eschotter/papers/Reichle_Schotter_2020_CogSciProc.pdf.

Reilly, R. G., & Radach, R. (2006). Some empirical tests of an interactive activation model of eye movement control in reading. *Cognitive Systems Research, 7*(1), 34–55. https://doi.org/10.1016/j.cogsys.2005.07.006

Saarela, M., & Georgieva, L. (2022). Robustness, stability, and fidelity of explanations for a deep skin cancer classification model. *Applied Sciences, 12*(19), 1–17. https://doi.org/10.3390/app12199545

Saksida, A., Iannuzzi, S., Bogliotti, C., Chaix, Y., Démonet, J. F., Bricout, L., & George, F. (2016). Phonological skills, visual attention span, and visual stress in developmental dyslexia. *Developmental Psychology, 52*(10), 1503. https://doi.org/10.1037/dev0000184

Schilling, H. E., Rayner, K., & Chumbley, J. I. (1998). Comparing naming, lexical decision, and eye fixation times: Word frequency effects and individual differences. *Memory & Cognition, 26*(6), 1270–1281. https://doi.org/10.3758/BF03201199

Schomacker, T., & Tropmann-Frick, M. (2021). Language representation models: An overview. *Entropy, 23*(11), 1422. https://doi.org/10.3390/e23111422

Seelig, S. A., Rabe, M. M., Malem-Shinitski, N., Risse, S., Reich, S., & Engbert, R. (2020). Bayesian parameter estimation for the SWIFT model of eye-movement control during reading. *Journal of Mathematical Psychology, 95*, Article 102313. https://doi.org/10.1016/j.jmp.2019.102313

Share, D. L. (1995). Phonological recoding and self-teaching: Sine qua non of reading acquisition. *Cognition, 55*(2), 151–218. https://doi.org/10.1016/0010-0277(94)00645-2

Share, D. L. (2008). Orthographic learning, phonological recoding, and self-teaching. In *Advances in child development and behavior* (Vol. 36, pp. 31-82). JAI.

Shultz, T. R. (2017). Constructive artificial neural-network models for cognitive development. In *(Eds)* N. Budwig, T. E., Zelazo, *New Perspectives on Human Development* (pp. 15-25), Cambridge University Press.

Sibley, D. E., & Kello, C. T. (2012). Learned Orthographic Representations Facilitates Large-Scale Modeling of Word Recognition. In *Visual Word Recognition Volume 1* (pp. 28-51). Psychology Press.

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. https://doi.org/10.1080/09541440903080583

Sibley, D. E., Kello, C. T., Plaut, D. C., & Elman, J. L. (2008). Large-scale modeling of wordform learning and representation. *Cognitive Science, 32*(4), 741–754. https://doi.org/10.1080/03640210802066964

Snell, J., & Grainger, J. (2019). Readers are parallel processors. *Trends in Cognitive Sciences, 23*(7), 537–546. https://doi.org/10.1016/j.tics.2019.04.006

Snell, J., & Grainger, J. (2019). Word position coding in reading is noisy. *Psychonomic Bulletin & Review, 26*, 609–615. https://doi.org/10.3758/s13423-019-01574-0

Snell, J., van Leipsig, S., Grainger, J., & Meeter, M. (2018). OB1-reader: A model of word recognition and eye movements in text reading. *Psychological review, 125*(6), 969. https://doi.org/10.1037/rev0000119

Staugaard, C. F., Petersen, A., & Vangkilde, S. (2016). Eccentricity effects in vision and attention. *Neuropsychologia, 92*, 69–78. https://doi.org/10.1016/j.neuropsychologia.2016.06.020

Vasilev, M. R., & Angele, B. (2017). Parafoveal preview effects from word N+ 1 and word N+ 2 during reading: A critical review and Bayesian meta-analysis. *Psychonomic Bulletin & Review, 24*, 666–689. https://doi.org/10.3758/s13423-016-1147-x

Veldre, A., Reichle, E. D., Yu, L., & Andrews, S. (2023). Lexical processing across the visual field. *Journal of Experimental Psychology: Human Perception and Performance, 49*(5), 649. https://doi.org/10.1037/xhp0001109

Weekes, B. S. (1997). Differential effects of number of letters on word and nonword naming latency. *The Quarterly Journal of Experimental Psychology Section A, 50*(2), 439–456. https://doi.org/10.1080/713755710

White, S. J., Drieghe, D., Liversedge, S. P., & Staub, A. (2018). The word frequency effect during sentence reading: A linear or nonlinear effect of log frequency? *Quarterly Journal of Experimental Psychology, 71*(1), 46–55. https://doi.org/10.1080/17470218.2016.1240813

Xiong, Y. Z., Qiao, C., & Legge, G. E. (2019). Reading with letter transpositions in central and peripheral vision. *Journal of Vision, 19*(3). https://doi.org/10.1167/19.3.17

Yang, L., Li, C., Li, X., Zhai, M., An, Q., Zhang, Y., & Weng, X. (2022). Prevalence of developmental dyslexia in primary school children: A systematic review and meta-analysis. *Brain Sciences, 12*(2), 240. https://doi.org/10.3390/brainsci12020240

Ziegler, J. C., Perry, C., & Zorzi, M. (2019). Modeling the variability of developmental dyslexia. In L. Verhoeven, C. Perfetti, & K. Pugh (Eds.), *Developmental dyslexia across languages and writing systems* (pp. 350–371). Cambridge University Press.

Zoccolotti, P., De Luca, M., Di Filippo, G., Judica, A., & Martelli, M. (2009). Reading development in an orthographically regular language: Effects of length, frequency, lexicality and global processing ability. *Reading and Writing, 22*(9), 1053–1079. https://doi.org/10.1007/s11145-008-9144-8

Xia, X., Liu, Q., Reichle, E. D., & Liu, Y. (2024). Saccadic targeting in the Landolt-C task: Implications for Chinese reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition.* https://doi.org/10.1037/xlm0001343