Letter Position Encoding in a Neural Framework

Ryan C. Stokes
University of California
Irvine, CA 92617
Email: stokesr@uci.edu

Gregory Hickok
University of California
Irvine, CA 92617
Email: greg.hickok@uci.edu

***Abstract*—Visual word recognition requires the identification of a word's component letters as well as their position within a word. The position of letters, once thought to be coded to the precise location in a word, is now understood to be more flexible. Indeed, a wealth of experimental evidence supports a view of visual word recognition that is forgiving to local changes in the position of letters. We present a model of letter position encoding, as an extension of the Overlap model, that treats letters as distributions along a normalized retinotopic space. Additionally, letters that do not belong in a target word, or are distant from the expected location, are modeled to inhibit the activation of the target word. This method has not yet been explored in the letter position encoding literature, and can help increase the biological plausibility of these models. We show that model estimates fit well with a database of priming studies used to investigate the effects of letter position manipulations on participants' reaction time and accuracy.**

I. Introduction

How the position of letters in a word are coded is a central question to models of visual word recognition. It is uncontroversial that some method of identifying letter position must exist, otherwise anagrams such as *angel* and *angle* would not be differentiable. An early solution, termed slot based coding, assigns each letter to its respective position. The word *SUNG* would be coded with an *S* in the first position, *U* in the second position, and so on. The arrangement would also be bound to the length of the word; the first letter in *SUNG* would not be comparable to the first letter in *SUN*. This was the assumption of the Interactive Activation (IA) model [1], [2], an influential model of visual word recognition and the predecessor of several other models [3], [4], [5], which share the assumption.

Slot based coding worked well for early models of visual word recognition, but it is unable to account for situations in which letters are removed, added, or scrambled with only a slight loss in readability. Word modifications such as these are often employed as experimental manipulations in studies of word recognition. Slot based coding can then be tested by comparing its predictions to the performance of human subjects. A condition might compare the difference in reaction time of a word such as *sung* to its close neighbor *snug*, a transposition of the middle two letters. A slot based approach assumes that only the first and last letter of *sung* will facilitate recognition of *snug*, comparable to the double replacement condition *sxxg*. Letters can also be inserted (e.g. sunxg), removed (e.g. sun), and replaced (e.g. sing) to create additional testable conditions. These types of letter manipulations can inform models of letter position encoding, and subsequently, models of visual word recognition. The following are several letter position effects that must be accounted for by any reasonable model of letter position encoding.

A. Transposition Effects

Transposed words are created by switching the place of two letters within a word. In all forms of slot based coding, this transposition would result in a 50% loss of information in a four letter word. However, as intuition tells us, and numerous experiments confirm [6], [7], [8], [9], there is a smaller impact on readability for words with transposed letters than words with letters that have been replaced randomly.

Transposition effects are most commonly tested with lexical decision tasks and same/different tasks using the masked prime paradigm developed by Forster & Davis [7]. In a masked prime paradigm, participants are presented with a brief (30ms–60ms) prime followed by a target. Participants are then instructed to either decide if the target is a word (lexical decision task) or choose between a target and a foil (same/different task). Early experiments using a 60-ms prime in a masked priming paradigm found significant transposition effects [8]. Transposition primes (anwser-ANSWER)[1] were used in a lexical decision task and compared to identity (answer-ANSWER) and single letter replacement primes (antwer-ANSWER). The authors reported a similar priming effect for transposition primes and identity primes, and a smaller effect for replacement primes [8].

Transposition effects have since become a gold standard for evaluating models of letter position encoding. The following accounts are some additional conditions that constrain transposition effects. First, increasing the distance between letters in a transposition causes more disruption to response times in a lexical decision task [9]. Additionally, conditions in which consonants and vowels were transposed displayed more disruption as compared to consonant-consonant or vowel-vowel transpositions [9]. Since most models of visual word recognition postulate two routes, one orthographic and the other phonological, we must consider whether the transposition effect can account for both. Finally, nonwords with typical transpositions (e.g. caniso-CASINO) were compared to nonwords with pseudohomophone transposed letter pairs

[1]Masked priming studies use lowercase primes and uppercase targets to control for visual shape similarities between the prime and the target.

978-1-5386-2726-6/17/$31.00

(e.g. kaniso-CASINO). The authors did not find an effect for the pseudohomophone pair and concluded that letter position encoding is primarily orthographic [10].

Several newer models of letter position encoding (e.g. Overlap model [11]; noisy channel model [12]) postulate a noisy input layer to account for transposition effects. If this is the case, then transposition effects should not be exclusive to letters. Transposition effects for digits [6], symbols [13], pseudoletters [13], and geometrical shapes [6] were evaluated with a procedure similar to the masked priming paradigm. Interestingly, digits, symbols, and geometrical shapes were all shown to produce robust transposition effects while pseudoletters were not [6], [13].

B. Relative Position Effects

Relative position effects pose an additional constraint on letter position encoding. A relative position stimulus can either be a subset, where some of the target letters are removed, or a superset, where irrelevant letters are inserted into the target stimulus. These modifications often violate the length of a target while maintaining relative letter position. For example, the word *apricot* can be reduced down to *arct* while still maintaining the relative letter positions compared to the full word. Relative position primes (e.g. arct) reliably facilitate target word (e.g. apricot) recognition both when length of the target is not violated (e.g. adricot [8]) and when the length is violated (e.g. arct [14]). The authors found that five letter primes would only facilitate a seven letter target if the letters did not violate the respective letter positions of the target. Surprisingly, adding hyphens between the letters so that absolute position information matched (e.g. a-r-c-t) did not increase the priming effect [14].

C. Letter Migration Effects

When two words are presented at the same time (e.g. cope and cage), the participant will occasionally respond with a new word that is composed of parts of each (e.g. cape). This is known as the letter migration effect as letters appear to migrate from one word to the other. The letters are not bound to the same relative position in the second word, and have been shown to span up to 2 letter positions [15]. For example, dividing a participant's attention between the words *step* and *soap* may elicit the response *stop*. If words were processed separately from one another these errors would not be produced. Instead, a shared space that processes words invariant to their position on the retina can merge word forms together, and process them concurrently.

D. Initial Letter Advantage

One last consideration involves the importance of the first letter, and to a lesser extent the last letter, of a word. Priming effects for transpositions are consistently higher for internal letters than for the initial or final letter positions [16]. This is consistent with the results of several experiments investigating the speed of letter in string processing [17], [18]. In one experiment, participants were briefly presented with a letter string followed by a pattern mask, then asked to identify the letter at a specified position. Letter accuracy tended to follow a W shape while speed of recognition followed an M shape [17]. The advantage of the initial and final position have been suggested to be due to crowding, as the initial and final position letters are only adjacent to one letter [19]. A crowding explanation does not, however, explain why symbols do not follow the same pattern as digits and letters [17]. A similar experiment had participants report on a same/different task between two letter strings which differ at one letter position. The researchers found higher accuracy and shorter reaction times for letter strings at the first position. This effect was prominent even when prime text is displayed at a random orientation (between horizontal and vertical). The authors suggest that the initial letter effect may arise from a rapid deployment of attention [18].

E. Towards a Model of Letter Position Encoding

Several models of letter position encoding can explain the aforementioned effects reasonably well. Prominent models include several variations of open-bigram models [20], [21], [22], [23], the Overlap model [11], and the Noisey Channel model [12]. Open-bigram models decompose words into unordered lists of letter pairs which can span several intervening letters. For example, the word *sung* would be coded as the unordered set: $\{$SU, SN, SG, UN, UG, NG$\}$. When presented with a transposed prime, such as *snug*, the resulting unordered bigram set: $\{$SN, SU, SG, NU, NG, UG$\}$, shares all but one bigram. This can be quantified as an 83% match, which would predict reasonable priming compared to the 50% match of the slot based method [20], [21]. Another approach to open bigram coding, the SERIOL (sequential encoding regulated by inputs to oscillations within letter units) model [22], [24], places bigrams on a continuous scale of activation. Letters are coded sequentially with a monotonically decreasing function of activation placing less emphasis on letters further from one another.

Open-bigram models can be effective at fitting general cases of transposition and relative position effects. However, several experiments have uncovered a lack of flexibility with these models [12], [25]. Bigram models share the assumption that letter pairs are the units of visual word recognition, but these units cannot be disturbed themselves. Nonetheless, a reversed bigram condition (e.g. ob-ABOLISH), which does not appear in the target's unordered bigram set, still produces priming effects [12]. The group also demonstrated that transpositions spanning more than two letter positions also produce priming effects [12]. Open bigram models typically limit the span to two letters, or in the case of Whitney's SERIOL model, have activation sharply drop off past two letters [24]. Finally, we consider an experiment that uses masked priming with first or last letter superset primes (e.g. wjudge- JUDGE and judgew-JUDGE respectively) and compared them to substitution-letter primes (e.g. juwge-JUDGE) [25]. An open bigram model predicts that the supersets, which contain the entire group of the prime bigram list, would prime the target more effectively.

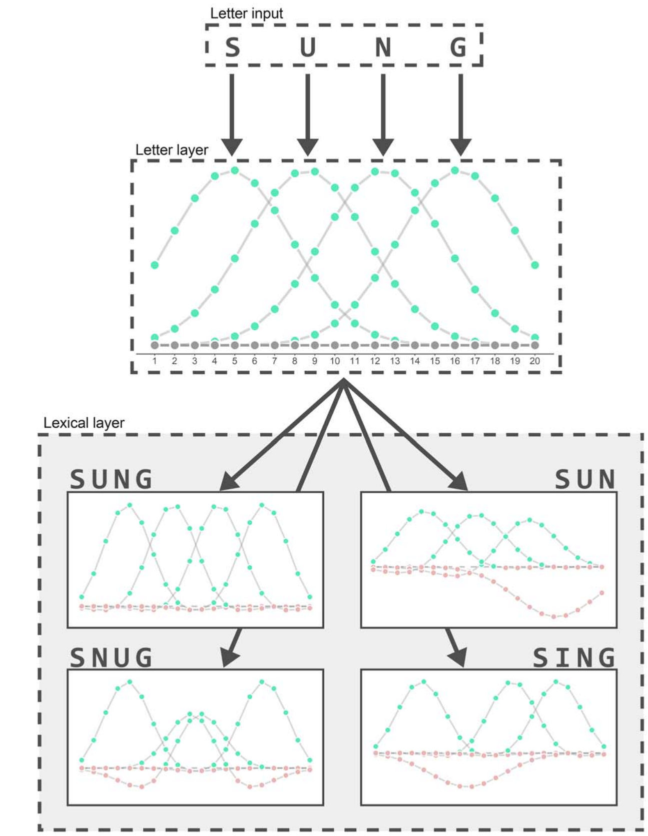

Fig. 1. An example of input into a model with four units in its lexicon: *SUNG, SUN, SNUG, SING*. The letter input represents the word as a set of ordered letters presented to the network. The letter layer converts these letters into distributions on a normalized retinotopic space. The amount of activation of each lexical item is then determined by the product of the letter layer's activation and the lexical weights.

The researchers instead found the first letter superset prime to be the most disruptive [25].

One concern of open bigram models is the need for two layers to code letter position. The first layer, a perfect representation of the position of each letter, is used to create a noisy replica of the same information. Some researchers question the rationale of coding a perfect representation of a word, then inducing a noisy filter to make it less precise [11]. The authors introduce a more parsimonious solution, termed the Overlap model, which posits that the initial letter position encoding is noisy, eliminating the need for a second layer. Rather than precise points, letters are understood better as distributions across the span of a word. This is similar to the SERIOL model, but with individual letters in place of bigrams. In this manner, the letter *a* in the word *trail* will activate the third position maximally, the second and fourth position to a lesser extent, and the first and fifth position minimally. Transposition effects are accounted for within the Overlap model since adjacent letters have high activation rates within several letter positions. Switching two letters only slightly reduces total activation for the target word [11].

II. Methods

In this paper we present a model of visual word recognition as an extension of the Overlap model, using a feedforward, rate coded, neural network. A fundamental assumption of the Overlap model, one that is shared with this model, is that the location of visual objects exist along a distribution in space, as opposed to specific points [11]. In the Overlap model, this space is divided into an appropriate number of slots depending on the length of the word, and the area under the curves of each slot are multiplied. The products are then summed across all slots resulting in a total activation level for a target word, given some prime. Our model does not divide space into slots for comparison, which assumes some knowledge about the discrete length of a word. Instead, the entire length of the retinotopic space informs the lexical layer. Additionally, our model adds inhibitory connections to suppress the activation of lexical entries with inserted or misplaced letters.

A. Model Specification

Two layers are used in this network, a letter layer and a lexical layer. The letter layer contains an array of N spatial units by 26 letters of the alphabet. Letter position is modeled as a graded response across a retinotopic map of neurons predicting an increasing activation with close proximity to the location of the letter (See Figure 1). A discretized Normal distribution (f_k) is used to represent the gradient of activation for each letter presented to the input layer. Activation of the letter layer at each unit *n* and at each letter *i* is calculated by generating the expected activation levels for each letter in the word at each position:

$$V_{ij} = \max_k (f_k(x_j | \frac{k}{L+1}, \sigma^2) I_{(A_i=\lambda_k)}) \tag{1}$$

Where $i = 1, 2, ..., 26$ and $j = 1, 2, ..., N$ index the cell array of the input layer, λ is the set of letters in the word presented to the network, k is the index of λ, A is the set of all letters in the alphabet, L is the length of the word, and $I_{(A_i=\lambda_k)}$ is an indicator variable to ensure only letters in the visual layer activate their respective neurons. x_i, the position of the i^{th} x between 0 and 1, is defined as:

$$x_j = \frac{j-1}{N-1} \tag{2}$$

A word presented to the network (e.g. $\lambda = \{W, O, R, D\}$) will activate all rows of the cell array where $A_i = \lambda_k$. The position of each letter in the word will determine the mean of the Normal distribution by dividing the index k by the length of the word plus one. If the same letter is repeated in a word then the maximum activation level between the two letters is used for each neuron. See Figure 2c for an example of a six letter word (e.g. $\lambda = \{L, E, T, T, E, R\}$) with the same letter in the middle two positions.

The network is trained by computing the weights between the letter layer and the lexical layer. Weights are set by first generating letter layer activations, then subtracting a constant α from each cell:

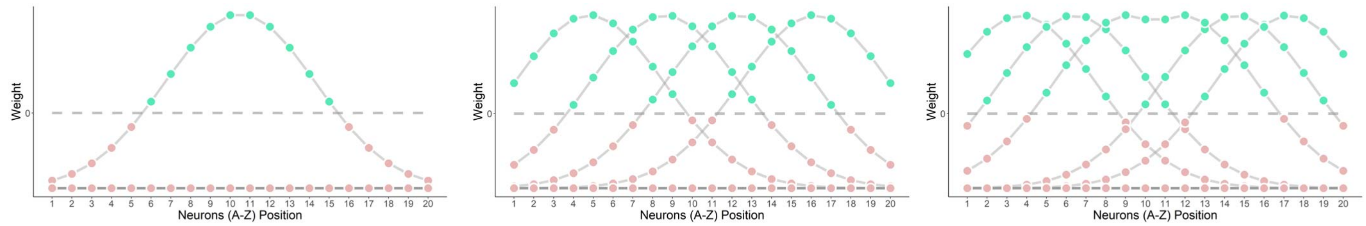

Fig. 2. Examples of lexical layer weights for when a single letter (a) 4 letter words (b) and 6 letter words (c) are trained to the network. Every word in the lexicon has a unique combination of 26 X 20 weights accounting for 26 letters across a retinotopic array of 20 neurons. When a word is composed of multiple of the same letter, the center two letters of (c), the maximum activation between the two is used. Letters which are not present in the word, or outside of the range of identification are given negative (inhibitory) weights.

$$W_{ij} = V_{ij} - \alpha \tag{3}$$

All letter rows that do not appear in the word start with a negative, inhibitory, weight. Additionally, spatial neurons that belong to a letter in the word can still receive inhibitory weights if they are far enough from the true position of the letter. The standard deviation of the letters remain constant, but the position of the letters, and the amount by which they overlap, will be determined by how many letters are in a presented word (Figure 2).

After training, the network was tested by presenting words to its input layer. Activation for lexical entries was found by summing over all input layer neurons multiplied by their respective weights and dividing by the number of letters in the input word:

$$T = \frac{1}{L} \sum_{i=1}^{26} \sum_{j=1}^{N} W_{ij} V_{ij} \tag{4}$$

Where T is the target word in a lexicon. See Figure 1 for examples of the activation levels of four lexical entries after the letter layer is presented with the word *SUNG*. Naturally, the lexical unit with the highest activation is the identity *SUNG*. The subset prime *SUN* is weighted to expect three letters in the center of the retinotopic space. These three letters still receive activation, but less so then if the letters were aligned. Additionally, the letter *G* inhibits activation since it is not a part of the expected word. The normalized activation space allows for strings of letters of any length to be compared with one another. In the case of transpositions, where two letters within a word switch places, the overlapping letters will still activate, but with less magnitude than if they were in the trained location (e.g. SNUG). Finally, a letter replacement such as *SING* will maximally excite letters in the appropriate location, but inhibit the letter that does not belong in the lexical entry[2].

[2]In fact *SING* also primes *SUNG* as they are close semantic neighbors [16]. However, the present model does not make semantic predictions so only the orthographic match is discussed here.

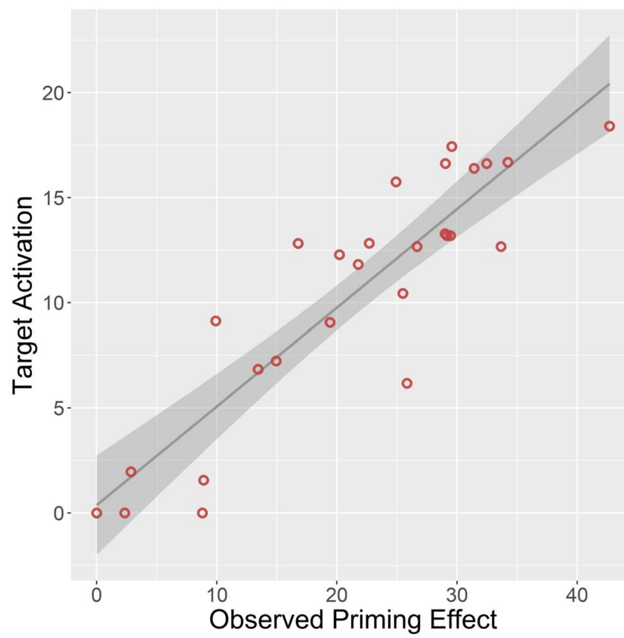

Fig. 3. Regression (95% confidence region) between the values estimated by the model on the Y axis and the observed relative reaction times of each condition on the X axis.

III. Results

The most common way to evaluate letter position effects is through masked priming. Experimental evidence routinely finds patterns of activation across a range of different priming conditions. However, the small effect sizes typically reported with masked priming studies, the inconsistencies with experimental design, and the differences in the population of participants makes comparing results across experiments problematic. To avoid these concerns, our model was instead tested using evidence from a masked priming mega study [26]. Megastudies tend to be more carefully controlled, have a higher number of participants, and most importantly, use the same methodology for all conditions. This results in more reliable priming estimates which are better suited for a qualitative comparison to our model's estimates. The study used 28 different modifications of 6 letter words as their primes,

TABLE I

Condition	Code	Adelman (2014)[a]	Current model[b]
Identity	123456	42.69	92.02
Final deletion	12345	34.23	83.41
Suffix	123456d	33.66	63.35
Final transposition	123465	32.46	83.11
Medial transposition	132456/124356/ 123546	31.42	81.97
Medial deletion	13456/12456/ 12356/12346	29.56	87.18
Final substitution	12345d	29.45	65.94
Initial substitution	d23456	29.16	65.94
Initial transposition	213456	29.03	83.11
Central insertion	123d456	29.00	66.40
Prefix	d123456	26.67	63.35
Half	123/456	25.83	30.87
Repeated central letter	123DD456	25.48	52.22
Central double deletion	1256	24.91	78.74
Medial substitution	1d3456/12d456/ 123d56/1234d6	22.68	64.14
Neighbor-once -removed	12d356/13d456/ 124d56/123d46	21.77	59.10
2-apart transposition	143256/125436	20.20	61.42
Central double insertion	123dd456	19.42	45.31
All-transposed	214365	16.77	64.11
Central double substitution	12dd56	14.94	36.19
Reversed halves	321654	13.44	34.20
3-apart transposition	153426	9.91	45.62
Interleaved halves	415263	8.90	7.80
Transposed halves	456123	8.80	0.00
Reversed-except-initial	165432	2.86	9.80
Central quadruple substitution	1dddd6	2.34	0.00
Unrelated	dddddd	0.00	0.00

[a] in milliseconds
[b] in arbitrary units

corresponding to many of the prime conditions used in the literature. The pseudoword condition was removed since our model does not distinguish pseudowords apart from random letter strings. The remaining 27 conditions were compared to our model using correlation (see Table 1 for the set of conditions used from [26]).

A. Masked Priming Mega Study

Values for each condition were generated by activating the letter layer with the same primes used in the mega study, then calculating the appropriate lexical entry's activation level. The number of cells in the letter layer cell array were held constant at 20 spatial units by 26 letter units. Two free parameters, σ^2 from (1) and α from (3), were sampled from a grid search to find good fits to the data. Activation levels from the model for 27 conditions have a correlation with experimental reaction times of r(25) = .886, $p < .01$ (See Table 1 for the predicted values of each condition), when $\sigma^2 = 0.7$ and $\alpha = 1.48$. Note, however, that the values generated by the model are activation levels and not an estimate of the expected reaction times.

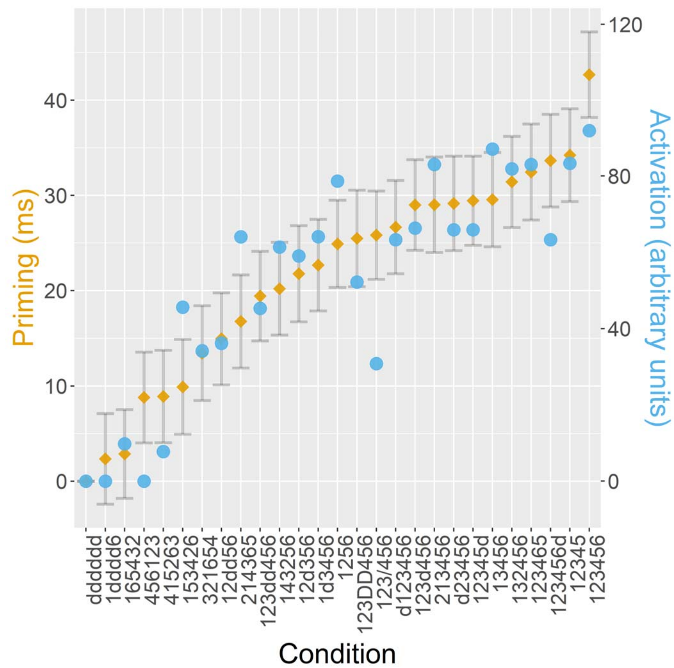

Fig. 4. The relative reaction time primes from the masked priming database (Y axis left) compared to the model's activation level for each condition (Y axis right).

With two free parameters the model displays a strong linear relationship with the masked priming data (See Figure 3). The relative activation levels of each condition compared against the reaction times of the masked priming data is also useful in determining the performance of the model (See Figure 4). The half word condition, where only the first or last three letters of a word act as the prime, does not correlate well with the masked prime data. It appears that more evidence is needed to make comparisons across words of different lengths as this is the only condition which compares a target and a prime which have size differences of more than two letters.

IV. Discussion

Our model of letter position encoding is informed by the various phenomenon reported by masked priming studies, and also constrained by the limitations of biological plausibility. Transposition effects, relative position effects, letter inhibition, and letter migration effects are each explained by the model using a simple network of neurons and two free parameters.

Transposition effects are explained at the letter layer. Adjacent letters have overlapping letter distributions that maintain some level of activation if those letters are transposed. However, distant letter transpositions will have inhibitory effects on word activation. This is because letter identification follows an activation gradient with the most activation centered around the true position of the letter, and inhibitory weights set near the tails. The model's prediction is consistent with behavioral studies where transpositions of more than two letter spaces have small priming effects on target words [26], [12].

Words with letters that have been removed (subset) or inserted (superset) will still generate reliable priming effects [26]. A shared space must then exist between words of different lengths for which comparisons can be made. The model accomplishes this by mapping every word onto the same normalized space, independent of the number of neurons in the letter layer. This strategy appears to be effective up to two letters of removal or insertion as the model underestimated performance of the half word (123/456) condition.

Letter migration effects can be informally evaluated by allowing the network to accept inputs from multiple letters at the same position. Two words can then be presented to the network at the same time, similar to an experiment where a word is presented to either side of the screen. This activates both sets of letters at overlapping positions. As expected, the lexical units for each of the two words will increase in activation, but occasionally, words which contain a combination of letters from the first two words will also activate strongly. The model is naturally flexible to letter position between words as well, eliciting strong activation for *stop* when presented with *step* and *soap* together.

Currently, the model does well to fit commonly studied masked priming effects. However, there are several simplifications that prevent the network from explaining specific types of conditions. For instance, the model has nothing to say about the initial letter position effect, where the first letter will be identified more quickly and accurately than other letters during a forced choice task. It is straightforward to include an additional parameter to increase activation of the initial letter, but this approach seems simplistic and unnatural. Instead, a model of attention that is integrated into the network is necessary to identify a word shape, and increase activation to the appropriate letter positions. This extension would fit more accurately with current theories of initial letter position effects [18]. Additional improvements, such as phonology, semantics, and syllable boundaries can also improve the explanatory power of the network and will likely be included in further iterations. The current model can most effectively be used as a front-end of letter position encoding in more advanced models of visual word recognition.

REFERENCES

[1] J. L. McClelland and D. E. Rumelhart, "An interactive activation model of context effects in letter perception: I. an account of basic findings." *Psychological review*, vol. 88, no. 5, p. 375, 1981.

[2] D. E. Rumelhart and J. L. McClelland, "An interactive activation model of context effects in letter perception: Ii. the contextual enhancement effect and some tests and extensions of the model." *Psychological review*, vol. 89, no. 1, p. 60, 1982.

[3] M. Coltheart, K. Rastle, C. Perry, R. Langdon, and J. Ziegler, "Drc: a dual route cascaded model of visual word recognition and reading aloud." *Psychological review*, vol. 108, no. 1, p. 204, 2001.

[4] C. Perry, J. C. Ziegler, and M. Zorzi, "Nested incremental modeling in the development of computational theories: the cdp+ model of reading aloud." *Psychological review*, vol. 114, no. 2, p. 273, 2007.

[5] M. S. Seidenberg and J. L. McClelland, "A distributed, developmental model of word recognition and naming." *Psychological review*, vol. 96, no. 4, p. 523, 1989.

[6] J. García-Orza, M. Perea, and A. Estudillo, "Masked transposition effects for simple versus complex nonalphanumeric objects," *Attention, Perception, & Psychophysics*, vol. 73, no. 8, pp. 2573–2582, 2011.

[7] K. I. Forster and C. Davis, "Repetition priming and frequency attenuation in lexical access." *Journal of experimental psychology: Learning, Memory, and Cognition*, vol. 10, no. 4, p. 680, 1984.

[8] K. I. Forster, C. Davis, C. Schoknecht, and R. Carter, "Masked priming with graphemically related forms: Repetition or partial activation?" *The Quarterly Journal of Experimental Psychology*, vol. 39, no. 2, pp. 211–251, 1987.

[9] H. I. Blythe, R. L. Johnson, S. P. Liversedge, and K. Rayner, "Reading transposed text: effects of transposed letter distance and consonant-vowel status on eye movements," *Attention, Perception, & Psychophysics*, vol. 76, no. 8, pp. 2424–2440, 2014.

[10] J. Acha and M. Perea, "Does kaniso activate casino?" *Experimental Psychology*, 2010.

[11] P. Gomez, R. Ratcliff, and M. Perea, "The overlap model: a model of letter position coding." *Psychological review*, vol. 115, no. 3, p. 577, 2008.

[12] D. Norris and S. Kinoshita, "Reading through a noisy channel: Why there's nothing special about the perception of orthography." *Psychological Review*, vol. 119, no. 3, p. 517, 2012.

[13] J. García-Orza, M. Perea, and S. Muñoz, "Are transposition effects specific to letters?" *The Quarterly Journal of Experimental Psychology*, vol. 63, no. 8, pp. 1603–1618, 2010.

[14] J. Grainger, J.-P. Granier, F. Farioli, E. Van Assche, and W. J. van Heuven, "Letter position information and printed word perception: the relative-position priming constraint." *Journal of Experimental Psychology: Human Perception and Performance*, vol. 32, no. 4, p. 865, 2006.

[15] C. J. Davis and J. S. Bowers, "What do letter migration errors reveal about letter position coding in visual word recognition?" *Journal of Experimental Psychology: Human Perception and Performance*, vol. 30, no. 5, p. 923, 2004.

[16] M. Perea and S. J. Lupker, "Does jugde activate court? transposed-letter similarity effects in masked associative priming," *Memory & Cognition*, vol. 31, no. 6, pp. 829–841, 2003.

[17] I. Tydgat and J. Grainger, "Serial position effects in the identification of letters, digits, and symbols." *Journal of Experimental Psychology: Human Perception and Performance*, vol. 35, no. 2, p. 480, 2009.

[18] A. J. Aschenbrenner, D. A. Balota, A. J. Weigand, M. Scaltritti, and D. Besner, "The first letter position effect in visual word recognition: The role of spatial attention." *Journal of Experimental Psychology: Human Perception and Performance*, vol. 43, no. 4, p. 700, 2017.

[19] D. G. Pelli and K. A. Tillman, "The uncrowded window of object recognition," *Nature neuroscience*, vol. 11, no. 10, pp. 1129–1135, 2008.

[20] J. Grainger and C. Whitney, "Does the huamn mnid raed wrods as a wlohe?" *Trends in cognitive sciences*, vol. 8, no. 2, pp. 58–59, 2004.

[21] J. Grainger and W. J. Van Heuven, "Modeling letter position coding in printed word perception." *Mental lexicon: "Some words to talk about words"*, pp. 1–23, 2004.

[22] C. Whitney, "How the brain encodes the order of letters in a printed word: The seriol model and selective literature review," *Psychonomic Bulletin & Review*, vol. 8, no. 2, pp. 221–243, 2001.

[23] S. Dehaene and L. Cohen, "The unique role of the visual word form area in reading," *Trends in cognitive sciences*, vol. 15, no. 6, pp. 254–262, 2011.

[24] C. Whitney, "Supporting the serial in the seriol model," *Language and Cognitive Processes*, vol. 23, no. 6, pp. 824–865, 2008.

[25] S. J. Lupker, Y. J. Zhang, J. R. Perry, and C. J. Davis, "Superset versus substitution-letter priming: An evaluation of open-bigram models." *Journal of Experimental Psychology: Human Perception and Performance*, vol. 41, no. 1, p. 138, 2015.

[26] J. S. Adelman, R. L. Johnson, S. F. McCormick, M. McKague, S. Kinoshita, J. S. Bowers, J. R. Perry, S. J. Lupker, K. I. Forster, M. J. Cortese *et al.*, "A behavioral database for masked form priming," *Behavior research methods*, vol. 46, no. 4, pp. 1052–1067, 2014.