

# Letter-position encoding and dyslexia

Carol Whitney

University of Maryland, USA

Piers Cornelissen

University of Newcastle, UK

This article focuses on applying the SERIOL model of orthographic processing to dyslexia. The model is extended to include a phonological route and reading acquisition. We propose that the temporal alignment of serial orthographic and phonological representations is a key aspect of learning to read, driving the formation of a phonemic encoding. The phonemic encoding and the serial representations are mutually reinforcing, leading to automatic, proficient processing of letter strings. A breakdown in any component of this system leads to the failure to form string-specific phonological and visual representations, resulting in impaired reading ability.

Following the pioneering work of Liberman and colleagues (1974), research into dyslexia has focused on phonological deficits. As discussed by Castles and Colheart (2004), there is a wide range of evidence that dyslexics are impaired in phonological awareness tasks, such as phoneme deletion, phoneme counting and phoneme lending. This correlation has largely been taken to reflect causality. That is, impaired phonological awareness is thought to reflect abnormal phonological representations, which are thought to be the fundamental cause of dyslexia. However, a causal relationship between phonological awareness and reading ability has not directly been established (Castles & Colheart, 2004). It may instead be the case that poor performance on phonological tasks is a result of poor reading ability. Indeed, several studies have shown that phoneme awareness tasks are influenced by properties of the corresponding orthographic representations (Castles et al., 2003; Stuart, 1990; Treiman & Cassar, 1997), indicating that performance on phonological tasks may tap into letter-based, rather than purely phonemic, representations.

Thus, deficits of phonological awareness may be a symptom, rather than a direct cause, of dyslexia; it may be more accurate to say that dyslexia results from the failure to form normal grapheme-phoneme associations (Castles & Colheart, 2004). Thus such a failure could potentially arise from the grapheme (visual) side, or from problems in forming the relevant visual-auditory associations. A deficit in any aspect of processing could then have repercussions throughout the entire network that is normally recruited to sub-serve reading.

To study dyslexia, it is useful to determine when and where processing first diverges from that of normal readers. MEG technology has revealed that this divergence occurs quite early, at the visual level. In normal subjects, a left-hemisphere (LH) infero-temporal

area was preferentially activated by letter strings (as opposed to symbols or faces) at about 150 ms post-stimulus (Tarkiainen, Cornelissen & Salmelin, 2002; Tarkiainen et al., 1999), while 80% of the dyslexic subjects did not show this activation (Helenius et al., 1999). Thus it seems that normal readers have learned some type of string-specific visual processing, while most abnormal readers have not.

What is the cause of this difference in visual processing? What can this difference tell us about the root cause(s) of dyslexia? To understand what goes wrong in visual processing in dyslexics, it is necessary to understand the nature of this processing in normal readers. What is the source of the early string-specific LH activation? What form do normal orthographic codes take? One approach to answering these questions is via computational modelling of how the early, retinotopic representation of a letter string is progressively converted into an abstract encoding of letter order. A recent model of letter-position encoding, dubbed the SERIOL model (Whitney, 2001; Whitney & Berndt, 1999), fits this bill. The SERIOL model was formulated via consideration of neurobiological constraints and behavioural data, and has yielded new insights into hemifield asymmetries in visual word recognition (Whitney, 2004a, 2004b; Whitney & Lavidor, 2004, 2005 forthcoming).

The purpose of this article is to provide an account of abnormal visual/orthographic processing in dyslexics, based on the SERIOL model. We consider how auditory and/or visual deficits could lead to a failure to learn normal string-specific visual processing, and how such a failure could contribute to difficulties in learning grapheme-phoneme correspondences. This discussion is a first step in applying the SERIOL model to dyslexia; it is admittedly quite sketchy and highly speculative. It is hoped that such consideration of the visual aspect of visual word recognition may inspire fruitful avenues of research in dyslexia.

This article is organised as follows. In the next section, the SERIOL model is presented. Then experimental evidence is presented supporting some key proposals of the model relevant to the issues at hand. These preliminary sections review previously completed work. We then extend the model to include a phonological route, and to address reading acquisition. After reviewing auditory and visual deficits in dyslexia, we are then in a position to apply the model to dyslexia. We discuss how visual/orthographic processing may break down during reading acquisition, and how this process may interact with phonological encoding.

### **Review of SERIOL model**

The SERIOL (Sequential Encoding Regulated by Inputs to Oscillations within Letter units) model is a theory of processing in the proficient reader. It specifies how an abstract letter position encoding is generated from a retinotopic representation, and how that encoding activates the lexical level. Note that SERIOL is not a full model of visual word recognition, as it does not address phonological processing. Rather, the model has focused on the orthographic route to the lexicon. However, this focus is not meant to imply a lesser role for phonological processing. A phonological route is assumed, but heretofore has not been elaborated. We will address the phonological route at a fairly high level later in this article.

We first give a brief overview of the existing model, and then discuss the processing in each layer in more depth. The model consists of five layers: edge, feature, letter, bigram and word. Each layer is comprised of nodes, which represent neural assemblies. Within

each layer, the activation of a letter is taken to be the total amount of activity across nodes representing that letter. To illustrate the representations and activation patterns at each layer, we will consider the stimulus CART, for fixation between the A and R.

The lowest layer of the model, the edge layer, corresponds to the early visual cortical areas, which are retinotopically mapped. At this level, there is an activation pattern arising from visual acuity. In our example, A and R would have the highest activation, and C and T would have lower activations (because there are more edge nodes representing A and R than C and T). At the feature layer, this acuity pattern is transformed into an activation gradient, dubbed the locational gradient, where activation decreases from left to right. That is, C would attain the highest activation, A the next highest, R the next and T the lowest activation. At the letter layer, the locational gradient interacts with oscillatory letter nodes to induce a serial encoding of order. That is, the C node fires, then A, then R, then T. At the next layer, temporally ordered pairs of letters activate bigram nodes. That is, a bigram node XY is activated when letter X fires before letter Y. In our example, bigram nodes \*C, CA, CR, AR, CT, AT, RT and T\* (where \* denotes a word boundary) become activated. The bigrams then activate lexical representations at the word layer. In our example, the word node CART would become more highly activated than any other word node.

Thus the retinotopic representation is converted into an abstract, location-invariant encoding of letter order via the creation of a serial representation. This serial encoding activates lexical representations via bigram nodes, which encode relationships between letters. See Figure 1 for a schematic of the letter through word layers. The choice of these representations and the transformations between representations are best illustrated in a top-down manner, as presented next.

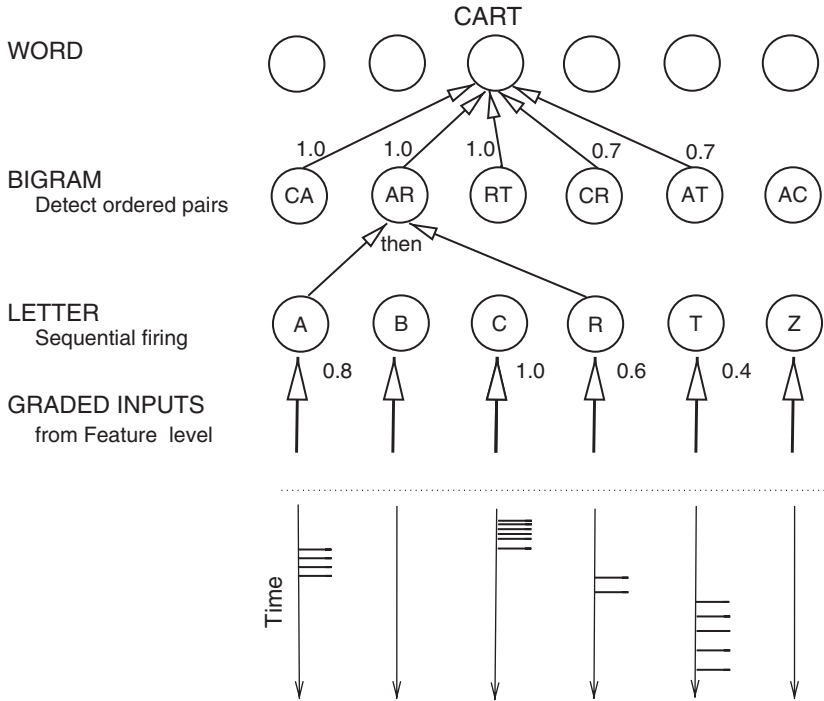
#### *Bigram layer to word layer*

First we consider what type of orthographic representation contacts the lexical layer. The results of priming experiments in which the order of letters from the target word is manipulated in the prime (Humphreys, Evett & Quinlan, 1990; Peressotti & Grainger, 1999) place strong constraints on the nature of the highest pre-lexical representation (Grainger & Whitney, 2004). These experiments indicate that order is important, not absolute position within the string. For example, GRDN primes GARDEN while GDRN does not, and the prime G\_RD\_N does not provide any more facilitation than GRDN. Thus, it is relative position that matters. These results can best be accommodated by units encoding the information that letter X preceded letter Y, where X and Y are not necessarily contiguous in the string.

The activation level of a bigram node is sensitive to the separation between the constituent letters. Activation decreases as the amount of time between the firing of the two letters increases. In our example, the RT node would have a higher activation than AT, which would have a higher activation than CT. Bigram nodes connect to word nodes via standard weighted connections, where the weights on the connections into each word node are proportional to the bigram activation pattern for that word.

#### *Letter layer to bigram layer*

How are such bigram nodes activated? Priming studies of letter trigrams indicate that priming can occur across letter positions (Peressotti & Grainger, 1995). This indicates that a letter detector can be activated by the corresponding letter in any string position.



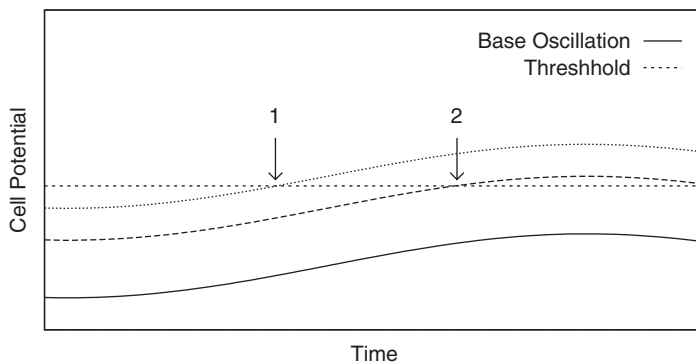
**Figure 1.** Architecture of the letter, bigram and word levels of the SERIAL model, with example of encoding the word CART.

*Notes:* At the letter level, simultaneous graded inputs are converted into serial firing, as indicated by the timing of firing displayed under the letter nodes. Bigram nodes recognise temporally ordered pairs of letters (connections shown for a single bigram). Bigram activations (shown above the nodes) decrease with increasing temporal separation of the constituent letters. Activation of word nodes is based on the conventional dot-product model.

Therefore, we assume that such position-independent letter nodes comprise the next lower layer of the model. Because these nodes only encode letter identity, positional information must be dynamically associated with such nodes. Two possibilities are that position is represented by activation level, or by firing order. Representation via activation level would require a monotonically decreasing activation gradient (e.g., in the CART example, the letter node C would have the highest activation, A the next highest, R the next and T the lowest). However, such an activation pattern at the letter layer is inconsistent with the well-known final-letter advantage; the final letter is perceived better than the internal letters, indicating a higher activation level than the internal letters. Therefore, in line with evidence for left-to-right string processing (Harcum & Nice, 1975; Nice & Harcum, 1976), letter order is taken to be represented serially. A bigram node is activated when its constituent letters fire in the correct order. For example, bigram node CA is activated when C fires before A, but not vice versa.

*Feature layer to letter layer*

How is this serial firing induced at the letter layer? Hopfield (1995) and Lisman and Idiart (1995) have proposed related mechanisms for precisely controlling timing of firing,



**Figure 2.** Interaction of input level and timing of firing for a cell undergoing a sub-threshold oscillation of excitability.

*Notes:* When a relatively high level of input (top curving line) is added to the base oscillation, the cell crosses threshold at time 1 (action potential not illustrated). If less input were received, the cell would cross threshold later in the cycle, such as at time 2.

in which nodes undergo synchronous, sub-threshold oscillations of excitability. For convenience, we designate the trough of this oscillatory cycle to be the 'start' of the cycle. Input level then determines how early in the cycle such a node can cross threshold and fire (see Figure 2). Near the beginning of the cycle, excitability is low, so only a node receiving a high level of input can cross threshold and fire. Excitability increases over time, allowing nodes receiving less and less input progressively to fire. Thus serial firing at the letter level can be accomplished via letter nodes that oscillate in synchrony and take input in the form of an activation gradient. In our example, the C node would receive the most input, A the next, R the next and T the least, allowing C to fire the earliest, A next, R next and finally T. Thus all nodes fire within a single oscillatory cycle, which is taken to be on the time scale of 200 ms (Lisman & Idiart, 1995).

An activated letter node inhibits other letter nodes. In our example, once C starts firing, how then does A ever start firing? As a letter node continues to fire, its firing rate slows, reducing lateral inhibition to the other nodes. This allows a new letter node to start firing. When an active letter node receives lateral inhibition, it then becomes strongly inhibited, so that it will not re-fire for the remainder of the oscillatory cycle.<sup>1</sup> Thus graded input levels and lateral inhibition create strictly sequential firing at the letter layer.

This process also creates varying activation levels. The activation of a letter node depends on both the rate and duration of firing. Under the assumptions that a higher input level leads to faster firing and that firing duration is fairly constant across letters, there is a decreasing activation gradient at the letter level. However, the node representing the final letter is not inhibited by a subsequent letter. It can continue to fire until the end (down-phase) of the oscillatory cycle.<sup>2</sup> Therefore, the final letter could potentially fire longer than the other letters, and reach a higher level of activation than the internal letters even though it receives less input. This is consistent with the well-known final-letter advantage. As discussed below, this proposal also explains some counter-intuitive experimental results on letter perceptibility.

Thus there must be a monotonically decreasing activation gradient across the next lower layer of the model, to provide input to the letter layer. Because this gradient

decreases from left to right (i.e. by spatial location), these lower-level units must be tuned to retinal location. Thus, a retinotopic representation is converted into a serial representation, creating a location-invariant encoding; location invariance is achieved by mapping space on to time. This location-invariant encoding is resumed to occur in the LH.

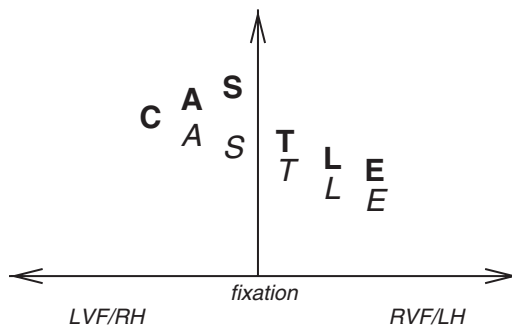
### *Edge layer to feature layer*

Based on the architecture of the visual system, there are several important characteristics of the edge layer that determine the nature of the transformations from the edge to the feature layer. The fibres from each retina divide, such that information reaching V1 is split by visual field, not by eye. The left visual field (LVF) projects to the right hemisphere (RH), while the right visual field (RVF) projects to the left hemisphere (LH). Available evidence indicates that there is little or no overlap in the cortical representation of the visual fields along the vertical meridian (Brysbart, 1994; Lavidor & Walsh, 2004; Leff, 2004). That is, letters immediately to the left of fixation are only projected to the RH, while letters immediately to the right of fixation are only projected to the LH. It is well known that the number of cortical cells representing a fixed area of space decreases as distance from fixation increases. Thus, activation decreases as eccentricity increases, giving a different activation pattern at the edge level from at the feature level (where activation decreases from left to right).

Therefore, the acuity pattern must be converted to the locational gradient as the edge layer activates the feature layer. Note that for a fixated word, the acuity pattern across the letters in the RVF/LH is the same as the locational gradient (i.e. decreasing from left to right). Thus the acuity gradient can serve as the locational gradient for those letters. However, in the LVF/RH, the acuity gradient increases from left to right; its slope is in the opposite direction as required for the locational gradient. Therefore, when the edge level activates the feature level, the acuity gradient must be inverted in the LVF/RH, while it can be maintained for the RVF/LH. Next we consider the details of this processing.

Obviously, activation levels cannot be increased by increasing the number of cells representing a letter. Rather, the locational gradient is created via modification of firing rates. In the LVF/RH, the acuity gradient is inverted via a combination of strong excitation and direction-specific lateral inhibition. This process is displayed in Figure 3. We propose that letter features in the LVF/RH become more highly activated by edge-layer inputs than those in the RVF/LH. This allows the first letter to reach a high level of activation. This could occur either via higher bottom-up connection weights from the edge layer, or by stronger self-excitatory connections within the feature layer. Within the RH feature layer, we propose that there is strong left-to-right lateral inhibition. That is, a feature node inhibits nodes to its right. As a result, letter features corresponding to the first letter receive no lateral inhibition, and inhibition increases as letter position increases. (Actually, there would be a gradient within each letter, but for simplicity we consider the mean activation level of a letter's features.) Thus, the features comprising the first letter attain the highest activation level (as a result of strong excitation and lack of lateral inhibition), and activation decreases towards fixation (due to sharply increasing lateral inhibition).

In the RVF/LH, the acuity gradient serves as the locational gradient. Overall excitation is weaker than to the LVF/RH. Left-to-right inhibition is not necessary, although some weak such inhibition may steepen the slope of the gradient. The two hemispheric



**Figure 3.** Formation of the locational gradient at the feature layer, for the centrally fixated stimulus CASTLE.

*Notes:* The horizontal axis represents retinal location, while the vertical axis represents activation level. The bold-face letters represent bottom-up input levels, which are higher in the RH than the LH. In each hemisphere, activation decreases as eccentricity increases, due to the acuity gradient. The italicised letters represent the effect left-to-right inhibition within the RH, and RH-to-LH inhibition in the LH. In the RH, C inhibits A, and C and A inhibition S, creating a decreasing gradient. The RH inhibits each letter in the LH by the same amount, bringing the activation of T lower than that of S. As a result, activation monotonically decreases from left to right.

gradients are ‘spliced’ together via cross-hemispheric inhibition. The RH features inhibit the LH features, bringing the activation of the LH features lower than the activation of the least activated RH features. As a result, an activation gradient that is strictly decreasing from left to right is created.

### *Summary*

The following are the important assumptions about processing at each layer.

#### *Edge layer*

- Retinotopic;
- representation of central vision split across hemispheres;
- activation levels based on acuity gradient.

#### *Feature layer (for a left-to-right language)*

- Retinotopic;
- representation still split across hemispheres;
- activation decreases from left to right (locational gradient);
- locational gradient formed by hemisphere-specific processing:
  - stronger excitation to RH than LH;
  - strong left-to-right lateral inhibition within RH;
  - RH inhibits LH.

#### *Letter layer*

- Position-independent letter nodes, located in LH;
- letter nodes undergo synchronous, sub-threshold oscillations in excitability;

- lateral inhibition between letter nodes;
- interaction of oscillations, lateral inhibition and locational-gradient input give a serial firing;
- letter-node activation level depends on:
  - firing rate – determined by input level (from locational gradient);
  - firing duration – determined by when next letter starts to fire, which is determined by the input level to that letter.

#### *Bigram layer*

- Bigram XY activated when letter X fires and then letter Y fires;
- activation of bigram XY decreases with the amount of time between the firing of letter X and letter Y.

#### *Word layer*

- Receives weighted connections from bigram layer.

### **Evidence for SERIOL model**

Having discussed the details of and motivations for the model, we now review some experimental support for the particulars of the model, based on novel experiments and new analyses of previous studies.

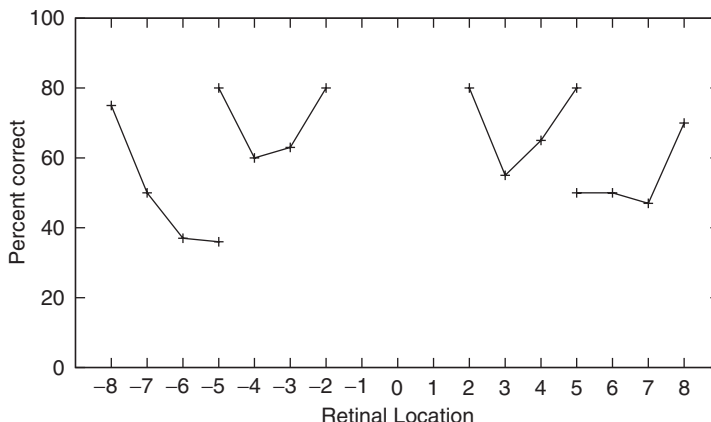
#### *Letter perceptibility*

It is well known that the external letters of a string (first and last letters) are the best perceived under central fixation. However, for unilateral presentation of short strings (three or four letters) at large eccentricities, a counter-intuitive pattern arises. In the LVF/RH, the first letter is the best perceived of all the letters; in the RVF/LH, the last letter is the best perceived (Bouma, 1973; Estes, Allemeyer & Reder, 1976; Legge, Mansfield & Chung, 2001). Thus, in each visual field, the letter farthest from fixation (where acuity is the lowest) is the most likely to be correctly reported. This pattern is present even at long exposure durations. For example, see Figure 4.

As discussed above, the induction of the serial encoding leads to differing activations at the letter level. These activation patterns depend on the interaction of the locational gradient and the oscillatory cycle. Such dynamics explain these patterns of letter perceptibility, as follows.

For a centrally fixated string, the initial-letter advantage and final-letter advantage arise for different reasons. The initial letter has an advantage because it receives the highest level of bottom-up input, allowing it to fire the fastest. It receives the most input because it is not inhibited from the left at the feature level. The final letter has an advantage because it is not inhibited by a subsequent letter during the induction of serial firing. That is, it is not inhibited from the right at the letter level. Thus, like others, we also attribute the advantage for the external letters to a lack of lateral inhibition. However, this reduced lateral inhibition does not arise from a lack of masking at a very low level (as is generally assumed). Rather it arises from string-specific processing, consistent with the finding that non-letter, non-digit symbols do not display an outer symbol advantage. That is, the outer symbol is the least well perceived, as would be expected on the basis of acuity (Mason, 1982).





**Figure 4.** Experimental results from Estes and colleagues (1974) for four-letter strings, occurring at two different retinal locations in each visual field.

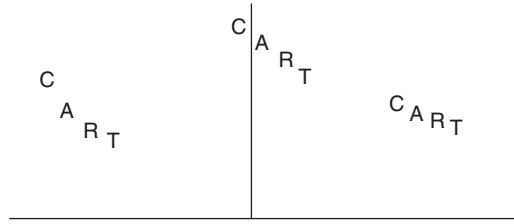
*Notes:* Exposure duration was 2400 ms. Subjects were trained to maintain central fixation, and their gaze was monitored.

This analysis of the external letter advantage implies that it should be possible to affect differentially the initial- and final-letter advantages. The initial-letter advantage should disappear if the amount of bottom-up input to the initial letter node is not significantly higher than to the other letters. The final-letter advantage should disappear if the final letter node starts firing late in the oscillatory cycle, and so is unable to fire for a longer time than the other letters. As we shall see, these proposals explain the counter-intuitive perceptibility patterns for lateralised presentation of short strings. First however, a more in-depth consideration of activation patterns at the feature level is required.

Recall that locational gradient formation requires different processing across the hemispheres. In the RVF/LH, the acuity gradient serves as the locational gradient. In the LVF/RH, the acuity gradient is inverted via strong bottom-up excitation and left-to-right lateral inhibition. Because the locational gradient is formed by different mechanisms in each hemisphere, the shape of the resulting gradient may vary with hemisphere, especially at large eccentricities. The acuity gradient is known to fall off less quickly as distance from fixation increases. That is, the slope of the acuity gradient is steepest near fixation, and becomes shallower as eccentricity increases. Because the RVF/LH locational gradient is based on the acuity gradient, this implies that the RVF/LH locational gradient becomes more shallow as eccentricity increases. (See right half of Figure 5.)

In the LVF/RH, formation of the locational gradient depends on left-to-right lateral inhibition. For strings at large eccentricities, inhibition may be too strong at early string positions, because of their relatively low level of activation (but, as discussed below, inhibition may become too weak at later string positions, because of the increasing acuity). (See left half of Figure 5.) Thus the prediction is that the locational gradient should vary with visual field. The proposal of a steeper LVF locational gradient (across string positions 1 to 4) explains an observed interaction between string position, eccentricity and visual field (Wolford & Hollingsworth, 1974), as discussed in Whitney (2001).

Now we are in a position to explain the unilateral perceptibility patterns. In particular, we will consider the results of Estes, Allemeyer and Reder (1976), given in Figure 4. In the



**Figure 5.** Schematic of locational gradients for the stimulus CART at three different presentation locations.

*Notes:* The vertical axis represents activation, while the horizontal axis represents retinal location. For central presentation, the gradient is smoothly and rapidly decreasing. For RVF presentation, the gradient is shallower because the acuity gradient is shallower. For LVF presentation, the initial letter strongly inhibits nearby letters, but the gradient flattens out as acuity increases.

following, primacy will signify that a letter is perceived better than all other letters, whereas advantage will mean that an external letter is perceived better than the internal letters.

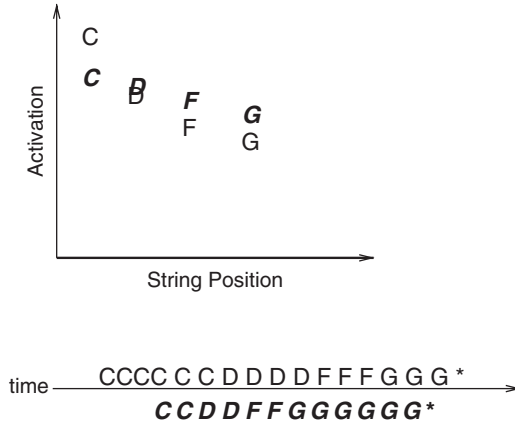
First we consider LVF presentation. At the feature layer, there is strong left-right inhibition, causing a steep locational gradient. Therefore, at the letter level, the first letter can fire for a (relatively) long time, as it is not quickly cut off by the next letter. Thus there is an initial-letter primacy. As a result of its low input level, the firing of the final letter is pushed late into the oscillatory cycle. Therefore, it cannot fire longer than the other letters, and no final-letter advantage emerges. This explains the perceptibility pattern for locations  $-8$  to  $-5$ .

For RVF presentation, there is weak left-to-right inhibition, while the acuity/locational gradient is quite shallow. Therefore the activation of the second letter's features is quite close to that of the first letter. As a result, at the letter level, the firing of the first letter is rapidly cut off by the second letter, giving no initial-letter advantage. Each successive letter quickly inhibits the preceding letter, allowing the final letter to start firing early in the oscillatory cycle. Therefore the final letter can fire longer than the other letters, creating a final-letter primacy. This explains the perceptibility patterns for locations 5 to 8. The proposed activation patterns for both visual fields are displayed in Figure 6.

This account explains the initial/final difference within a single retinal location (at  $-5$  and 5 in Figure 4). In the LVF/RH, the left-to-right, feature-level inhibition creates a disadvantage for a final letter, whereas an initial letter does not receive this inhibition. In the RVF/LH, the shallow locational gradient creates a disadvantage for an initial letter because its firing at the letter layer is rapidly inhibited by the second letter. For a final letter, firing at the letter layer can continue until the end of the oscillatory cycle instead.

In contrast to the patterns at the larger eccentricity, the perceptibility function is U-shaped for both  $-5$  to  $-2$  and 2 to 5. As a result of higher acuity, bottom-up input is higher overall. In the LVF/RH, this allows the final letter to start firing earlier in the cycle, creating a final-letter advantage. Along with the usual initial-letter advantage, this gives the U-shaped pattern. In the RVF/LH, the acuity/locational gradient is steeper than for the larger eccentricity, so the difference in input to the first and second letters is larger, creating an initial-letter advantage and giving an overall U-shape.

Next we consider the implications of this account for differing exposure durations. Under the assumption that a longer exposure duration increases the overall level of bottom-up input, the above analysis suggests that the RVF final-letter primacy and the LVF initial-letter primacy should be differentially affected by variations in durations. In



**Figure 6.** Locational gradients and resulting firing patterns for L VF/RH presentation (normal font) and R VF/LH presentation (bold italics).

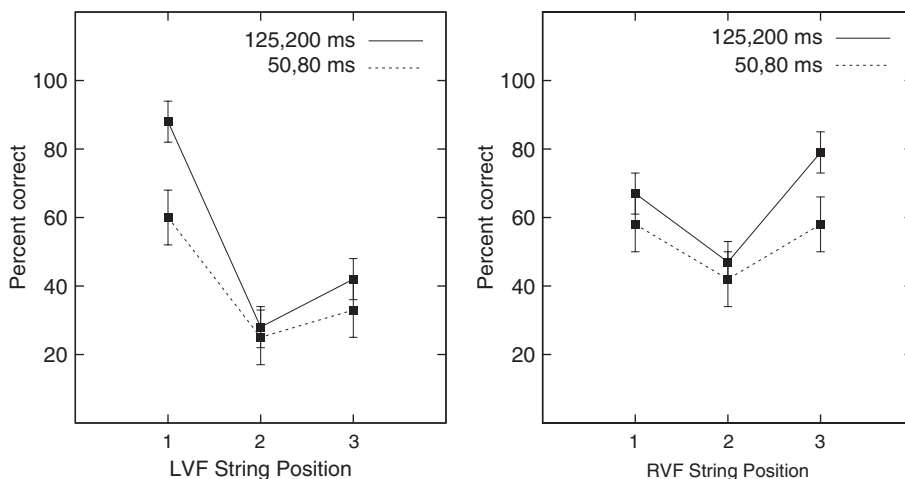
*Notes:* Top: Comparison of locational gradient for string CDFG under R VF/LH presentation and L VF/RH presentation. Bottom: Cartoon of resulting firing pattern at the letter level. The point in the oscillatory cycle at which the down phase prevents further firing is marked \*. In the L VF/RH, the first letter fires faster and longer than the other letters, because it receives a much higher level of input. The variations in the amount of bottom-up input create decreasing activation across the string. The final letter starts firing late in the cycle, and is soon cut off by the end of the oscillatory cycle, giving no final-letter advantage. In the R VF/LH, each letter rapidly cuts off the previous letter, allowing the final letter to fire a long time. As a result, activation is flat across the string and rises for the final letter. These firing patterns account for the perceptibility patterns at the larger eccentricities in Figure 4.

the R VF, we would not expect to see a final-letter primacy at very brief exposures, because the very low level of input pushes the firing of the final letter late into the oscillatory cycle. As exposure duration increases, the firing of all the letters is shifted earlier and earlier into the cycle, allowing the final letter to fire longer and longer. In contrast, the activation of a non-final letter should not change much, because its firing is still quickly cut off by the subsequent letter. Thus, in the R VF, a final-letter primacy should emerge as exposure duration increases.

However, in the L VF, the initial-letter primacy should be present at very brief durations, because strong left-to-right inhibition at the feature level does not depend on temporality. As exposure duration increases, the initial letter should be the primary beneficiary because, at the feature level, the increased bottom-up input to non-initial letters is cancelled by increased lateral inhibition from the first letter.

To summarise, in the R VF, the final-letter primacy should not be present at very brief exposures. Increasing exposure duration should primarily benefit the final letter, creating a final-letter primacy. In the L VF, the initial-letter primacy should be present at very brief exposures. Increasing exposure duration should primarily benefit the initial letter, increasing its primacy.

A search of the literature revealed that a relevant experiment had already been performed, in which retinal location and exposure duration were systematically varied in a trigram identification task (Legge, Mansfield & Chung, 2001). However, the published data were not presented in a way that would allow evaluation of the above predictions. So the first author requested the raw data from the authors, who kindly provided it. The data



**Figure 7.** Results from Experiment 2 (Legge et al., 2001) for the two largest eccentricities, grouped by exposure duration, with 95% confidence intervals.

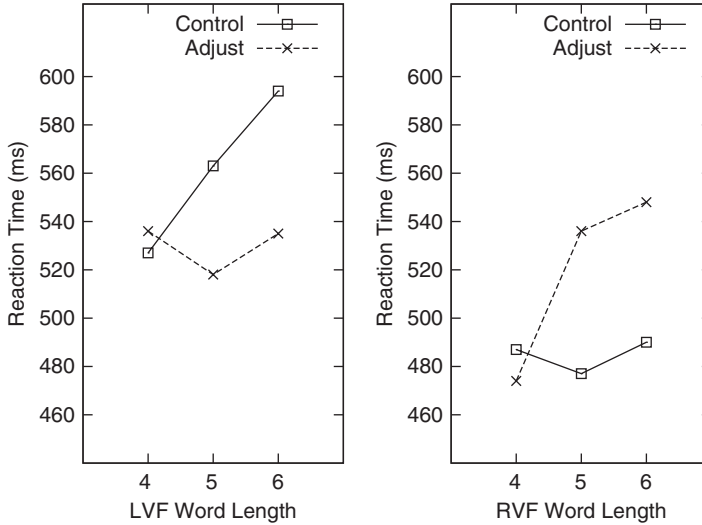
were analysed for the largest two eccentricities ( $-11$  to  $-9$  and  $-10$  to  $-8$  versus  $8$  to  $10$  and  $9$  to  $11$ , in units of letter widths) for very brief exposures ( $50$  ms and  $80$  ms) versus longer exposures ( $125$  ms and  $200$  ms). This analysis did indeed reveal the predicted patterns, as shown in Figure 7 (Whitney, forthcoming).

### *Lexical asymmetries*

Different patterns for the left and right visual fields have also been observed for lexical properties, such as length. For the lexical-decision task, length has no effect on reaction times (RTs) for RVF/LH presentation, while each additional letter increases RT by  $20$ – $30$  ms for LVF/RH presentation (Ellis, Young & Anderson, 1988). This asymmetry has generally been taken to reflect dual modes of lexical access, with inefficient, letter-by-letter activation of the lexicon for the LVF/RH, and efficient, parallel activation for the RVF/LH (Ellis, Young & Anderson, 1988), although the source of the asymmetry has remained a subject of debate (Jordan, Patching & Thomas, 2003; Nazir, 2003).

Indeed, the dual-modes account is inconsistent with brain-imaging evidence indicating that lexical access is routed through the LH, independently of presentation location (Cohen et al., 2002). Instead, this asymmetry may reflect the asymmetric activation patterns discussed above. For a long word in the LVF/RH, the resulting locational gradient would not be smoothly decreasing. Strong inhibition to the second and third letters would make their activations quite low, while inhibition may be insufficient at letters close to fixation. As a result, the locational gradient is initially steeply decreasing, and then flattens out. Such an activation pattern would provide a non-optimal encoding of letter order, which could increase settling time at the word level, thereby increasing reaction time. Thus a length effect may occur in the LVF/RH because the locational gradient becomes more and more degraded as string length increases.

This analysis suggests that it should be possible to abolish the LVF length effect by making the locational gradient smoother. Activation level can be experimentally manipulated by changing contrast level. Thus, for a six-letter word, increasing the



**Figure 8.** Results from lexical decision experiment for word targets.

*Notes:* Under the Adjust condition, contrast levels were manipulated as described in the text.

contrast of the second and third letters, while decreasing the contrast of the final letter should result in a smoother gradient, thereby abolishing the length effect. That is, reaction times to LVF six-letter words under such a contrast adjustment should be equal to four-letter words under normal presentation. However, in the RVF, such a contrast adjustment should degrade a previously smooth gradient, creating a length effect. We performed such a study (Whitney & Lavidor, 2004), and these predictions were confirmed, as shown in Figure 8. These results demonstrate that a length effect is not an inherent feature of RH processing, suggesting that the dual-modes theory is incorrect. Rather, the asymmetry of the length effect is due to activation patterns, as predicted and explained by the SERIOL model. Whitney and Lavidor (forthcoming) have also shown that activation patterns are the source of a hemifield asymmetry related to the effect of orthographic neighbourhood size (Lavidor & Ellis, 2002).

However, if letter-level activations occur serially, as proposed, it should take more time for all the letters of a long word to fire than of a short word. Why then is there no effect of string length for RVF (and central) presentation? It may be the case that settling time is increased for a short word (as compared to a long word). That is, the reduced amount of bottom-up input to the word layer (from fewer letters) could potentially increase the amount of time required to reach response criterion after the final letter fires. If this increased settling time for a shorter word were to offset exactly the earlier firing of its final letter, there would be no effect of length. Thus reaction times are not necessarily a reliable indicator of serial versus parallel processing.

A recent study demonstrates the need for these more complex scenarios. New and colleagues (New, Ferrand, Pallier & Brysbaert, forthcoming) undertook an investigation of the effect of word length based on the English Lexicon Project, which is an online database of lexical-decision RTs for over 40,000 words (Balota, Cortese, Sergent-Marshall & Spieler, 2004). Once the effects of frequency, number of syllables and orthographic-neighbourhood size were factored out, they found that RTs actually

decrease with increasing string length for words of three to five letters,<sup>3</sup> are constant with length for words of five to eight letters, and increase with length for words of eight or more letters. Thus string length has differing effects over different lengths. It is highly unlikely that these effects reflect differences in the method of lexical access. Rather, these results most likely indicate that the effect of length is the sum of multiple influences, where the relative strength of opposing components varies with length. For example, the data are explained by assuming a facilitatory component which decreases in strength as word length increases, coupled with a steady linearly increasing cost of string length due to a serial encoding. Such a facilitatory effect may arise from decreased settling time at the word level as a result of more bottom-up activation, and this effect may be strongest for short words and then taper off, reaching a ceiling level. Thus for short words, the facilitatory effect dominates; for medium-length words, the two components cancel each other out, as discussed above; for long words, the cost of seriality dominates.

### *Serial processing*

The most controversial aspect of the SERIOL model is the proposed serial representation of letter order. How could the proposal of seriality be investigated if reaction times do not provide a reliable index? In order to investigate a temporal phenomenon, time should be used directly. Harcum and Nice (1975) used this approach in a clever experiment in which pairs of eight-letter compound words were very briefly presented in sequence. The pairs were selected to allow meaningful blends. For example, the stimuli 'headache' and 'backrest' could be recombined to give 'headrest' or 'backache'. When fixating on the centre of the stimuli, subjects tended to report the first half of the first word, and the second half of the second word (e.g. 'headache' then 'backrest' yielded 'headrest'). This result unambiguously shows sequential readout. The first half of the first word was processed first. By the time that the second half of the stimulus was reached, the stimulus had changed and the second half of the second word was processed.

They also included trials where fixation fell within the first or second half of the stimuli. Fixation within the second half yielded the same response pattern as central fixation. However, for fixation within the first half of the stimuli, the pattern reversed (e.g. 'backache' tended to be reported instead of 'headrest'). The authors took these results as evidence for left-to-right processing for central fixation, and peripheral-to-central processing for non-central presentation. However, we propose a more parsimonious explanation, based entirely on left-to-right processing. It has been shown that fixation within the first half of a word provides the Optimal Viewing Position (OVP) and the fastest processing (O'Regan, LevySchoen, Pynte & Brugailiere, 1984). Therefore, when fixation fell at the OVP, there may have been time to process the first word in its entirety. Then the second word would have been processed starting at the beginning, overwriting the representation of the first word. The second word was presented more quickly than the first, so there may only have been enough time to process its first half. Therefore, the response was comprised of the first half of the second word, and the second half of the first word.

### *Summary*

These accounts provide explanations of letter perceptibility patterns and hemifield asymmetries that are otherwise difficult to explain. Thus they provide strong support for the theory of locational gradient formation at the feature level, and for the proposed

interaction of the locational gradient and the oscillatory cycle at the letter level. As discussed below, learning to form such a locational gradient would be a crucial aspect of reading acquisition. First, however, we discuss the proposed nature of the phonological route in visual word recognition. (In some dual-route terminologies this would be referred to as the sub-lexical as opposed to the lexical route).

### Phonological route

The reading system must exploit the visual and phonological processing used in object recognition and language. Therefore, we consider how the phonological route in visual word recognition may form on top of the phonological processing system. Of course, this first requires understanding the phonological processing system. Based on models proposed by Hickok and Poeppel (2004), and Levelt (2001), we propose the following architecture.

In the middle superior temporal gyrus, the auditory input is encoded into acoustic features suitable for speech recognition. This representation is then transformed into a phonological encoding in posterior superior temporal gyrus. This phonological representation contacts lexico-semantic representations in middle temporal gyrus.

What is the nature of this phonological representation? Experiments have indicated that the granularity of auditory word recognition is likely sub-phonemic. As a vowel's spectrum shifts due to the place of articulation of the upcoming consonant, this information about place of articulation is used to restrict possible lexical candidates even before the identity of the consonant is known (Warren & Marslen-Wilson, 1988). Further studies have indicated that this information is likely not relayed via partial activation of a phonemic level, but rather via direct contact with the lexical level. Thus lexical activation depends on abstract phonetic features (Marslen-Wilson & Warren, 1994).

How are the phonetic features matched to phonological word forms? Phonetic features are activated serially by the incoming auditory input. However, the long-term representation of word forms cannot be based on (serial) firing patterns, but rather must be static (i.e., based on connection weights). It is difficult to see how to compare directly a serial representation with a stored representation. Rather, it is more straightforward to assume that activated phonological representation depends on a parallel encoding of the phonetic relationships, allowing comparison of an activation pattern (vector) with a weight vector. This is not to say that the entire phonetic encoding must be present for such a comparison to be made. For example, recall that in the SERIOL model, bigrams are activated sequentially and lexical activations occur incrementally, yet the bigram encoding itself does not depend on serial firing. That is, all bigrams could be activated at the same time and they would still encode the same information; the encoding of the information itself does not directly depend on serial bigram activations, as is evident from a simulation in which bigrams were activated in parallel (Whitney & Berndt, 1999).

Thus we assume that there is a mechanism for converting serial, phonetic input into a parallel encoding of that sequence. We propose a similar solution as to the bigram case. That is, the parallel phonetic representation is based on open biphones, which encode that phonetic feature A occurred before phonetic feature B. Thus, the biphone encoding would activate lexical items via the same type of mechanism as the bigram encoding.

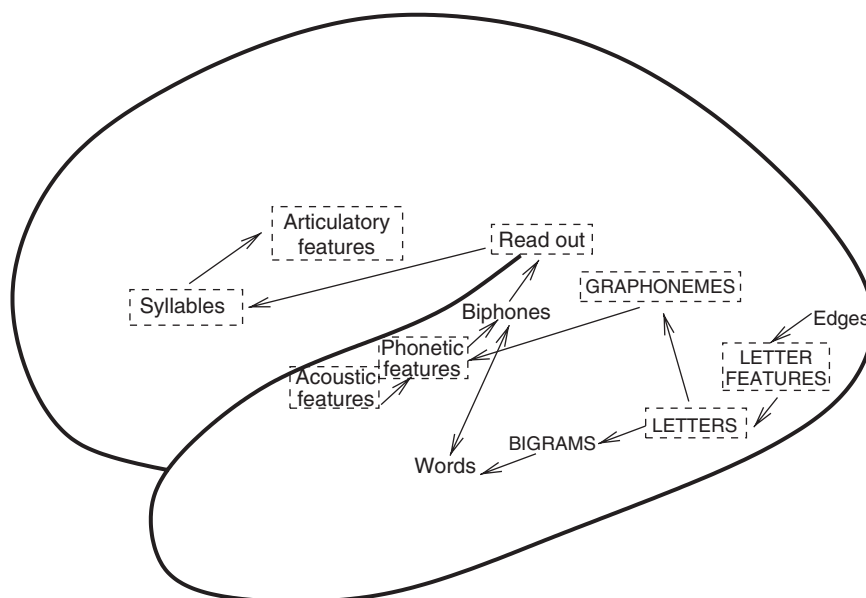
For output, this biphone representation is read out via areas near the temporo-parietal junction (Hickok & Poeppel, 2004), to be mapped to an articulatory representation in

inferior frontal cortex. As discussed by Levelt (2001), output is likely to be based on syllabic units, which are constructed on the fly. That is, the biphone representation is read out serially, and converted into a sequence of syllables.

We propose that preliterate phonological encodings do not include a level of representation corresponding to phonemes. (The appearance of phoneme-like units in categorical perception and in speech errors may arise from representation at a lower, biphone level.) Rather, reading acquisition itself creates a phonemic representation, via linkages of graphemes to groups of phonetic features. We will use the term ‘graphoneme’ to refer to such phonemes, highlighting the proposal that the phonemic encoding depends on a linkage to orthography. Thus the phonological route in reading activates graphonemes, then phonetic features and then biphones, allowing recognition along the auditory word-recognition pathway. Graphonemic representations would likely reside in the angular gyrus, a multi-modal association area at the occipital/temporal/parietal boundary. It is well known that lesions near this region cause alexia.

During reading, interactions between the orthographic and phonological routes would occur at both the word and biphone levels, as the phonological route activates lexical representations (via an assembled biphonic representation), and the orthographic route activates stored biphonic representations (via lexical items). The resulting biphonic representation would specify the ultimate output of reading the string. See Figure 9 for an overview of the proposed system.

We propose that phonological awareness tasks are performed over the graphonemic representation, consistent with evidence for orthographic influences in such tasks. Thus



**Figure 9.** Schematic of proposed reading and phonological system, where representations are shown approximately at their anatomical locations.

*Notes:* Dashed boxes indicate representations that depend directly on serial firing to encode order information. Capital letters indicate novel representations created by reading acquisition. Arrows show primary direction of processing, but are not meant to rule out processing in the other direction, or other pathways.



the inability of dyslexics to perform phonological-awareness tasks would indicate the absence of graphonemic representations. This proposal is consistent with the results of brain imaging studies of phonological tasks (Grunling et al., 2004). In normal readers, such tasks activate posterior brain areas; in dyslexics, such tasks activate more anterior, pre-frontal regions, likely indicating an attempt to perform such tasks over a syllabic representation.

This proposal is also consistent with the double dissociation in lexical/sub-lexical abilities observed across patients with focal lesions, as discussed in Hickok & Poeppel (2004). Some patients are unimpaired in auditory word recognition, yet cannot perform sub-lexical tasks involving syllable identification and/or phoneme manipulation, while some patients show the opposite pattern (ability to do sub-lexical tasks, but impairments in auditory word recognition). The former pattern would result from damage to graphonemic representations and/or syllabification procedures, while leaving the phonetic-feature/biphone/lexical pathway intact. Indeed, such a pattern results from lesions near the temporoparietal junction. The latter pattern would result from a disconnection between biphone and lexical representations. Indeed, such a pattern results from lesions to inferior temporal cortex. Overall, these patterns indicate that the sub-lexical representations used in meta-linguistic tasks are a super-set of those used in auditory word recognition, as we suggest.

For the present purposes, the specification of the phonological route at this high level is sufficient. We leave for future work the further (non-trivial) details of the neural encoding of the phonological representations.

### **Reading acquisition**

Thus far we have discussed the principles of the SERIOL model of orthographic processing, experimental evidence for those principles, and conjectures about the phonological route in visual word recognition. Next we consider which aspects of the proposed orthographic processing must be innate, and which are learned. The learned aspects then have implications for what goes wrong when visual and phonological processing of letter strings is impaired for dyslexics.

#### *What is innate and what is learned?*

The edge layer of the SERIOL model is based on known properties of the primary visual areas, and these properties are therefore innate.

At the feature layer, the left-to-right nature of the locational gradient is learned, as it depends on reading direction. (In a language read from right to left, the locational gradient would decrease from right to left.) Furthermore, distinguishing objects by horizontal relationships is unnatural. The identity of a natural object does not change as it is rotated around the vertical axis; a lion is still a lion regardless of whether it is facing to the left or to the right. Thus the visual system must learn to distinguish horizontal order for the purpose of processing words, and it must learn to impose a monotonically decreasing activation gradient.

At the letter layer, the serial encoding depends on the oscillatory nature of letter nodes. Obviously, the brain does not learn to use oscillations to encode information. Rather, the general mechanism of using oscillatory nodes to create a location-invariant representation

via the conversion of spatial representation (with varying activation levels) into a temporal representation is taken to be innate. For convenience in the remainder of this article, we will refer to the serial firing resulting from such an interaction as a scan. Such low-level scans do not rely directly on visual attention, but rather arise from the interaction of activation gradients and oscillatory cells. (However, visual attention could affect such a scan by modulating the way in which low-level activation patterns are filtered.) We assume that an innate oscillatory mechanism operates in the visual system during normal object recognition. For present purposes, this would take the form of a radial scan from fixation towards the periphery. Such an outward scan would depend directly on the acuity gradient. Thus the proposal is that formation of the locational gradient modifies the normal, symmetrical, outward processing to produce an asymmetrical left-to-right scan for string processing (in left-to-right language, of course). Indeed, Efron, Yund & Nichols (1990) have presented evidence for an outward visual scan, which seems to be perceptual and not attention based, and the pattern of the scan is influenced by the literacy of the subject (Ostrosky-Solis, Efron & Yund, 1991).

At the next layer of the model, bigram nodes represent the ordering between two letters. Thus, relationships between the sub-elements (letters) of the stimulus are encoded as a set of pairs. The general capacity to represent relationships in this way is taken to be innate. Thus in the case of strings, the visual system learns to form bigram nodes to represent relationships between pairs of letters.

The word level of the model corresponds to the lexicon. Obviously, people must learn to associate an orthographic representation with the corresponding lexical/semantic representations. Such learning would be based on general mechanisms of Hebbian/associationist learning.

Thus these general representational mechanisms are taken to be innate: the pairwise representation of relationships, the existence of oscillatory cells, and the capacity to use these oscillatory cells to convert space into time via differences in activation levels.

Processing that is highly specific to visual word recognition occurs primarily at the feature level. Recall that, in normal readers, there was early, string-specific, LH inferotemporal activation (Tarkiainen et al., 1999). This activation may be associated with the detection of a letter string, and the triggering of string-specific processing (i.e. induction of the locational gradient). It may be the case that dyslexics fail to learn this string-specific processing. Next we consider how normal readers may learn to form the locational gradient, and then we are in a position to discuss what can go wrong in dyslexic readers.

### *Learning to read*

Most early readers have been taught the sounds that letters make. Such conscious knowledge of letter sounds is likely to be based on lexico-semantic encodings. That is, knowledge that 'a B makes the /b uh/ sound' may be encoded similarly to the knowledge that 'a dog says /w uh f/', where /w uh f/ and /b uh/ are stored as phonological word forms associated with lexical items.

For simplicity in the following discussion, we will equate a letter and a grapheme. At the earliest stage of reading acquisition, we assume that the reader fixates each letter sequentially, focusing visual attention onto the fixated letter; this focused attention allows only that letter to become activated. If the reader is explicitly sounding out the word, each letter activates the lexical sound representation. The early reader may blend these sounds

together via the same mechanism that blends morphemes' phonological representations together to make a word.

Once a pronunciation of the word is available, the reader can fixate each letter in sequence, while slowly saying or rehearsing the word. Alternatively, a teacher may slowly say the word while pointing to the corresponding letter(s). Thus, the sequence of letters and the sequence of phonetic features are temporally aligned; each letter fires in close proximity to a corresponding phonetic representation. This temporal synchrony then provides the basis for the formation of intermediate representations that link the orthographic and phonetic representations (i.e. for the formation of graphonemes). Over time, as graphoneme units become more stable, the phonological representation can be directly formed via activation of a sequence of phonetic features, rather than by blending a sequence of lexical sound representations.

If the reader fixates on the first letter but fails to restrict attention to that letter, all letters will automatically rapidly fire in order, under the above assumption of an automatic outward visual scan. If graphonemes have been learned, this will allow the rapid generation of the corresponding phonetic representation. However, if graphonemes are not present, there will be insufficient time to consciously activate the lexical sound representation for each letter, and thus the reading system will not be able to take advantage of such rapid visual processing. Thus the formation of graphonemes also allows faster visual processing; sequential fixations are no longer necessary, as a word can be rapidly processed by fixating on the first letter.

However, fixation on the first letter limits the visibility of letters near the end of the word. The reader then learns to process a word with fixation falling nearer the centre. Consider a four-letter word. Assume that the reader fixates between the first and second letters, evenly distributing visual attention across all the letters. This strategy will lead to an incorrect graphonemic representation, because the outward visual scan would cause letters one and two to fire at the same time (then three, then four, because of the RVF/LH acuity gradient). Thus the generated graphonemic representation would not be serial and would be erroneous, and therefore the phonetic features and biphones would not be activated correctly.

However, if the reader happens to pay more attention to the letter on the left (the first letter) than the letters on the right, the first letter would fire before the second letter. (Think of a top-down attentional gradient, rather than a bottom-up locational gradient, across the feature level.) Then letters in positions three and four will fire. This would cause the corresponding graphonemes to be activated sequentially, providing a suitable phonetic representation. Therefore, the reader learns to invoke an attentional gradient across the letters to produce left-to-right, serial processing. As a result, the graphonemes learned during the first stage then reinforce sequential processing of letters during subsequent stages of learning.

So, initially there is a serial letter-based representation as a result of sequential fixations. When this serial representation is temporally aligned with the blended phonological representation, graphonemes are formed. Graphonemes then allow processing of multiple letters in a single fixation. As such a shift occurs, there is pressure to maintain a serial letter-based representation, in order to generate a serial phonetic representation via the graphonemes. The generated phonetic representation (or the phonetic representation of a correction provided by a teacher) can be internally replayed in synchrony with the rapid, serial letter-based representation. This temporal correspondence then further reinforces the formation of graphoneme units and automatic processing.

Over time, the top-down attentional gradient would drive learning of automatic, bottom-up formation of an activation gradient (i.e. the locational gradient). Indeed, a simulation utilising weight adjustments on both excitatory and inhibitory connections demonstrated the feasibility of such learning. The system was able to learn to convert a symmetrical activation gradient (corresponding to the acuity gradient) into a monotonically decreasing activation gradient for strings at varying ‘retinal locations’ (Whitney, 2004b, p. 95).

In summary, the key proposals are as follows.

- (1) The initial formation of graphoneme units is driven by the synchronous activation of a single letter and the corresponding phonetic features.
- (2) The graphoneme units and the serial nature of phonology drive serial processing of letters within a single fixation, which in turn reinforces the formation of graphoneme units.
- (3) The serial processing of letters within a single fixation is initially driven by a top-down attentional gradient. Over time, the visual system learns to form an activation gradient in a bottom-up manner (i.e. learns the proposed edge-to-feature level processing in the SERIOL model).

Consistent with the proposed importance of serial processing in learning to read, young readers show a strong length effect, which slowly diminishes in magnitude until adult competence is reached, at which point a length effect is absent (Aghababian & Nazir, 2000). Because the length effect steadily decreases, its disappearance is likely a result of increasing efficiency, rather than a shift from serial to parallel processing. That is, the per-letter processing time steadily decreases as readers become more proficient. At pre-adult competence levels, the per-letter processing time dominates, giving a length effect. At the adult level of proficiency, the per-letter processing time is short enough that the time cost of an additional letter is offset by speeded settling at the lexical level (due to increased bottom-up input from the additional letter), giving no length effect.

As above, we leave the details of this proposal for future research. In particular, mechanisms for the formation of graphoneme units are non-trivial. In a language like English, where there is no simple one-to-one correspondence between graphemes and phonemes, learning to create graphoneme units would be particularly challenging, consistent with the increased prevalence of dyslexia in English compared to languages with shallow orthographies (Spencer, 2000). However, any account of the learning of grapheme-phoneme correspondences must solve alignment/segmentation problems; a key proposal here is that temporality contributes to the solution of these problems.

### **What can go wrong?**

Next we review sensory/cognitive deficits observed in dyslexics in tasks not directly related to reading or phonological awareness. Based on the present account of reading acquisition, we discuss how such deficits could lead to abnormal processing of letter strings.

#### *Auditory deficits*

It is well known that dyslexics show deficits in phonological awareness tasks. But do they show differences from normal readers in more basic aspects of the processing of spoken

language? Behavioural and ERP studies have given evidence for categorical problems in phoneme detection (Mody, Studdert-Kennedy & Brady, 1977; Schulte-Korne, Deimel, Bartling & Remschmidt, 1998; Tallal, 1980). However, most such studies have used synthetic speech sounds; a recent study using both natural and synthetic speech showed categorical perception deficits in dyslexics for the synthetic speech only (Blomert & Mitterer, 2004), suggesting that perceptual auditory deficits do not interfere with the processing of ecologically valid speech input. An ERP study of auditory lexical decision, where the target was preceded by a prime, showed differences in the N1 and N2 time windows for normal versus dyslexic children (Bonte & Blomert, 2004). In particular, N1 amplitude was reduced in dyslexics, and was insensitive to alliteration between prime/target onsets, while normals' N1 amplitude was modulated by alliteration. In contrast, normals and dyslexics did not differ in the N400 time window. The authors conclude that the dyslexics have a deficit of early phonetic/phonological processing, but not of phonological/lexical processing.

But do such differences directly cause dyslexics' reading disability? By the definition of dyslexia, dyslexics have normal spoken language ability. Thus, such differences were not severe enough to cause difficulty with the automatic acquisition of spoken language. If auditory/phonological deficits did not interfere with such a complex, implicit learning task, how could they interfere with reading, which is explicitly and systematically taught?

We suggest that the observed deficits are not causal per se, but rather are concomitant symptoms of anatomical anomalies, where the root problem is abnormal left-hemisphere connectivity from auditory cortex into other cortical areas (Paulesu et al., 1996), such as the angular gyrus in particular. Indeed, diffusion tensor MRI has revealed micro-structural abnormalities in white matter of dyslexics' temporo-parietal region (Klingberg et al., 2000). PET studies have shown strong functional connectivity (activation covariance) between Wernicke's area, left angular gyrus and left extrastriate cortex in normals during pseudoword reading, but no functional connectivity between those areas in dyslexics (Horwitz, Rumsey & Donahue, 1998; Pugh et al., 2000). This lack of functional connectivity may well be a consequence of weakened anatomical connectivity.

Therefore, we propose that reading acquisition breaks down as follows. During the first stage (single-letter processing under sequential fixations), the formation of graphonemic representations is not as robust as usual. That is, the visual and phonological representations are essentially normal, but there is a failure adequately to form associations between them to create graphonemes.

As discussed above, a lack of graphonemes would limit the speed with which letters could be visually processed, preventing employment of the automatic visual scan. The continued co-activation of a letter and a lexically-based sound representation would reinforce this style of processing, further precluding the development of graphonemes. So some readers may remain at the stage of slow, explicit blending (yielding the so-called surface or dyseidetic subtype). Thus, letter processing is serial (but induced via sequential within-word fixations) and grapheme-phoneme correspondences are encoded via lexical sound representations. Sub-lexical processing occurs via successive loops through the lexical/orthographic route, where lexical access occurs at the letter level, rather than for the word as a whole.

Other readers may instead forgo serial processing. Under an absence of graphonemes, an outward visual scan would not create an erroneous phonetic representation; conversely, a left-to-right scan would not give a suitable phonetic representation. Therefore, there is no pressure to create an attentional gradient to support rapid, left-to-right processing of

letters. Rather the reader processes letter strings like other visual objects. This lack of a serial representation then further precludes the formation of graphonemic units. As a result, the reader adopts a holistic word-recognition strategy, with no sub-lexical processing (yielding the so-called phonological or dysphonetic sub-type). However, without a left-to-right scan, the encoding of letter order is not robust, so orthographic lexical access is not as reliable as in normal readers. The lack of a robust encoding of letter order would also contribute to spelling problems.

In both cases, there is a failure to learn the mechanisms for string-specific visual processing, which further contributes to abnormal phonological processing. Next we consider how a similar breakdown could occur from the visual side.

### *Visual deficits*

Recent research has revealed a magnocellular deficit in some dyslexics (Demb, Boynton & Heeger, 1998; Livingstone, Rosen, Drislane & Galaburda, 1991; Stein & Walsh, 1997). The dorsal stream of the visual system, which specialises in motion processing, spatial localisation and attention, primarily receives inputs from the magnocellular system (Maunsell, Nealey & DePriest, 1990). Interestingly, it has been shown that magnocellular function (as indexed by coherent-motion detection threshold) is correlated with letter-position encoding ability (as indexed by the ability, in lexical decision, to reject non-words that are anagrams of words) (Cornelissen, Hansen, Gilchrist, Cormack, Essex & Frankish, 1998). Thus it has been proposed that magnocellular dysfunction may lead to visual attentional difficulties which result in reduced ability to localise the position of letters in a string (Cornelissen et al., 1998; Vidyasagar, 2001, 2004).

Indeed, dyslexics exhibit abnormalities in visual-attention tasks. Some studies have shown LVF mini-neglect and RVF over-distractibility (Facoetti & Molteni, 2001; Hari, Renvall & Tanskanen, 2001). Others have shown deficits in serial (but not parallel) search (Iles, Walsh & Richardson, 2000), and a lack of a benefit for pre-cueing the location of a target (Roach & Hogben, 2004).

Thus there is evidence that some dyslexics are unable to focus visual attention normally. Recall that the proposed initial stage of reading acquisition depends on the ability to activate a single letter (per fixation) in tandem with hearing the corresponding phonetic features. If the early reader is unable to localise visual attention to the fixated letter, all letters will fire in each fixation, within the time span of each phoneme. Thus it will not be possible to form reliable graphoneme representations, because graphemes are temporally associated with non-corresponding phonetic features.

Without reliable graphoneme units, there is no pressure to invoke rapid, left-to-right processing of letters. Furthermore, attentional difficulties may preclude the formation of the top-down attentional gradient necessary to generate such processing via the low-level visual scan (unless fixation always falls on the initial letter).

The end result is the same as in the auditory-deficit case of non-serial processing. String-specific processing is not learned, and graphoneme representations are not formed. Words are processed like other visual objects, leading to a whole-word recognition strategy.

Vidyasagar (2001, 2004) has made a somewhat similar proposal, suggesting that proficient readers sequentially deploy attention across the string in a rapid, top-down manner, and that dyslexics are unable to do so. In contrast, we propose that proficient readers have learned to form an activation gradient in a bottom-up manner; this locational gradient automatically

yields serial processing. Learning to form this locational gradient depends on the ability to form a top-down attentional gradient; the impetus for forming this top-down attentional gradient depends on learned associations between letters and phonetic features; the formation of these learned associations initially depends on the ability to focus visual attention onto a single letter.

### Discussion

We have proposed that there are two important aspects to reading acquisition:

- (1) The creation of a serial phonemic representation that is not present for spoken language.
- (2) The induction of string-specific visual processing that operates from left to right.

These two novel types of serial representations mutually reinforce each other, allowing proficient reading. When the graphoneme representation does not develop normally, because of an inability to focus visual attention or to abnormal connectivity between the auditory and visual areas, the string-specific visual processing does not develop normally either. When string-specific visual processing does not develop normally, rapid phonetic processing is not possible.

In layman's terms, dyslexia is associated with inability to distinguish reversible letters, and with mis-ordering of letters. These symptoms are indeed present in dyslexic children at much higher rates than normal readers (Kaufman, 1980; Terepocki, Kruk & Willows, 2002), but they have been largely ignored of late in the scientific literature. A purely phonological approach to dyslexia does not explain the prevalence of such errors.

In contrast, consideration of the proposed visual aspect of processing does. Both types of reversals may be symptoms of the absence of the locational gradient to induce rapid, left-to-right processing. Note that for an outward visual scan (where feature activation levels depend directly on acuity in both VFs), the relative activation levels of the features of 'd' vary with fixation. When a 'd' is in the LVF, the straight vertical segment has a higher activation level than the curved segment. When a 'd' is in the RVF, the curved segment has a higher activation level than the straight segment. Thus it is not possible to form an invariant recognition procedure of 'd' (versus 'b') based on the activation levels of the constituent features. However, under the locational gradient, the curved segment of 'd' always has a higher activation level than the straight segment (and vice versa for 'b'), allowing such an invariant recognition procedure. Thus we suggest that reliable identification of reversible letters depends on the presence of left-to-right processing. If the normal encoding of letter order depends on such an automatic left-to-right scan, the absence of such a scan will impair the encoding of letter order, explaining the concurrence of the two types of reversal errors. Furthermore, if the encoding of letter order is impaired, the normal procedures for mapping grapheme sequences to phoneme sequences will be disrupted.

How could aspects of the present proposals be tested? As discussed above, we have proposed that counter-intuitive letter perceptibility patterns (particularly for short strings at large eccentricities) arise from the formation of the locational gradient and the conversion of this activation gradient into serial firing. If this processing is absent in dyslexics, they should show a different letter perceptibility pattern – one that is more dependent on acuity.

What are the implications for treatment of dyslexia? Continuing with the speculative nature of this work, we offer some suggestions. We propose that phonological training is successful insofar as it promotes formation of non-lexically based phonemic units. For example, the task 'remove the N sound from /l a n d/' would promote processing of phonetic sequences on the single phoneme level, without requiring the use of lexicalised phonemic representations (and so should be beneficial). In contrast, the task 'divide /l a n d/ into its phonemes' would require that the output be given in lexicalised phonemic representations (and so would not be as beneficial).

The association between letters and phonemes should be reinforced implicitly via temporal synchrony between visual and phonological representations. For example, a word could be displayed on a computer monitor, and, as the pronunciation of the word is provided, the corresponding letters are sequentially brightened. (There should then be some task related to the word or to the letters, in order to promote vigilance during the sound/letter display). This strategy should promote the formation of direct linkages between letters and phoneme units that have been established by phonological training, cementing the creation of graphonemes.

Then further visual training would be important, in order to promote rapid serial processing. This could be provided by flashing a fixation point on the screen, and then flashing a short word, where the task is to read the word. The location of the fixation point would be varied across trials, as would the position of the word relative to the fixation point. This would provide training in the allocation of visual attention, and in multi-letter processing within a single fixation. Moreover, there should be a contrast gradient across the letters. This would mimic a top-down attentional gradient and promote left-to-right processing via the automatic visual scan. Hopefully, such a contrast gradient would foster learning of bottom-up formation of the locational gradient. Over time, the contrast gradient could be phased out.

A previous study has shown that treatment utilising brief presentation (100 to 300 ms) of words, either centrally or randomly lateralised, improved spelling ability in dyslexics, whereas longer central presentation (1500 ms) or presentation to a single visual field did not (Lorusso, Facoetti & Molteni, 2004). This increased spelling ability may reflect a more reliable orthographic lexicon, stemming from more robust letter-position encoding. Perhaps brief presentation alone forced formation of the locational gradient, because overt scanning was not an option. It would be interesting to see if imposition of a contrast gradient would generate a greater improvement in reading/spelling ability than standard stimuli.

The suggested treatments would be appropriate for any root cause(s) of dyslexia. The hope is that such training would result in the formation of graphonemes and string-specific visual processing. Once such representations were firmly established, reading acquisition might then occur normally.

In conclusion, letter strings are unique visual objects because they are comprised of symbols representing sounds. This induces a left/right asymmetry that is not present for other objects. Because of these factors, letter strings require specialised visual processing. The SERIOL model provides a theory of this processing, indicating that learning to form a rapid, serial encoding of letter order may be an indispensable aspect of becoming a proficient reader. This proposal of a serial encoding was originally motivated by the necessity of providing a mechanism to create a location-invariant representation of letter order. Moreover, such a serial encoding meshes well with the serial nature of phonology. To fully understand what goes wrong in dyslexia, it would seem necessary to understand



normal visual processing in reading, and how it interacts with phonological processing. The present work is an initial step in this direction.

### Acknowledgements

The final form of this paper derives from an extensive and detailed e-mail dialogue between the two authors during the course of Summer and Autumn 2004. Piers Cornelissen would like to acknowledge that a considerable part of his contribution to this dialogue stems from his long-standing collaboration with Dr Peter Hansen at the University Laboratory of Physiology, Oxford. We would also like to thank the two anonymous reviewers for their inspiring comments.

### Notes

1. This raises the question of how repeated letters are handled. We assume that there are multiple copies of each letter mode, and a different mode becomes activated for each instance.
2. This assumes that a single word is being processed, as in experimental studies. Under natural reading conditions, multiple short words could be represented in a single oscillatory cycle.
3. It is likely that the reason that this facilitatory effect of word length has not been previously observed is that the effect of orthographic-neighbourhood size (N) was not controlled. N is the number of words that can be formed by changing one letter of the target to another letter (Coltheart, Davelaar, Jonasson & Besner, 1977). High N is actually facilitatory for words in lexical decision (Andrews, 1997). Because N generally decreases with word length, reduced N facilitation for longer words may have cancelled out the facilitatory effect of more letters.

### References

- Aghababian, V. & Nazir, T.A. (2000). Developing normal reading skills: Aspects of the visual processes underlying word recognition. *Journal of Experimental Child Psychology*, 76, 123–150.
- Andrews, S. (1997). The effect of orthographic similarity on lexical retrieval: Resolving neighborhood conflicts. *Psychonomic Bulletin and Review*, 4, 439–461.
- Balota, D., Cortese, M., Sergent-Marshall, S., Spieler, D.H. & Yap, M. (2004). The English Lexicon Project: A web-based repository of descriptive and behavioral measures for 40,481 English words and nonwords. Available at <http://elexicon.wustl.edu>. Washington University.
- Blomert, L. & Mitterer, H. (2004). The fragile nature of the speech-perception deficit in dyslexia: Natural vs synthetic speech. *Brain and Language*, 98, 21–26.
- Bonte, M.L. & Blomert, L. (2004). Developmental dyslexia: ERP correlates of anomalous phonological processing during spoken word recognition. *Cognitive Brain Research*, 21, 360–376.
- Bouma, H. (1973). Visual interference in the parafoveal recognition of initial and final letters of words. *Vision Research*, 13, 767–782.
- Brybaert, M. (1994). Interhemispheric transfer and the processing of foveally presented stimuli. *Behavioral Brain Research*, 64, 151–161.
- Castles, A. & Coltheart, M. (2004). Is there a causal link from phonological awareness to success in learning to read? *Cognition*, 91, 77–111.
- Castles, A., Holmes, V.M., Neath, J. & Kinoshita, S. (2003). How does orthographic knowledge influence performance on phonological awareness tasks? *Quarterly Journal of Experimental Psychology A*, 56, 445–467.
- Cohen, L., Lehericy, S., Chochon, F., Lemer, C., Rivaud, S. & Dehaene, S. (2002). Language-specific tuning of visual cortex? Functional properties of the Visual Word Form Area. *Brain*, 125, 1054–1069.
- Coltheart, M., Davelaar, E., Jonasson, J. & Besner, D. (1977). Access to the internal lexicon. In S. Dornic (Ed.), *Attention and Performance VI*. (pp. 535–555). Hillsdale, NJ: Lawrence Erlbaum Associates.

- Cornelissen, P.L., Hansen, P.C., Gilchrist, I., Cormack, F., Essex, J. & Frankish, C. (1998). Coherent motion detection and letter position encoding. *Vision Research*, 38, 2181–2191.
- Demb, J.B., Boynton, G.M. & Heeger, D.J. (1998). Functional magnetic resonance imaging of early visual pathways in dyslexia. *Journal of Neuroscience*, 18, 6939–6951.
- Efron, R., Yund, E.W. & Nichols, D.R. (1990). Detectability as a function of target location: Effects of spatial configuration. *Brain and Cognition*, 12, 102–116.
- Ellis, A.W., Young, A.W. & Anderson, C. (1988). Modes of word recognition in the left and right cerebral hemispheres. *Brain and Language*, 35, 254–273.
- Estes, W., Allemeier, D. & Reder, S. (1976). Serial position functions for letter identification at brief and extended exposure durations. *Perception and Psychophysics*, 19, 1–15.
- Facoetti, A. & Molteni, M. (2001). The gradient of visual attention in developmental dyslexia. *Neuropsychologia*, 39, 352–357.
- Grainger, J. & Whitney, C. (2004). Does the human mind read words as a whole? *Trends in Cognitive Science*, 8, 58–59.
- Grunling, C., Ligges, M., Huonker, R., Klingert, M., Mentzel, H.J., Rzanny, R., Kaiser, W.A., Witte, H. & Blanz, B. (2004). Dyslexia: The possible benefit of multimodal integration of fMRI- and EEG-data. *Journal of Neural Transmission*, 111, 951–969.
- Harcum, E.R. & Nice, D.S. (1975). Serial processing shown by mutual masking of icons. *Perception and Motor Skills*, 40, 399–408.
- Hari, R., Renvall, H. & Tanskanen, T. (2001). Left minineglect in dyslexic adults. *Brain*, 124, 1373–1380.
- Helenius, P., Tarkiainen, A., Cornelissen, P., Hansen, P.C. & Salmelin, R. (1999). Dissociation of normal feature analysis and deficient processing of letter-strings in dyslexic adults. *Cerebral Cortex*, 9, 476–483.
- Hickok, G. & Poeppel, D. (2004). Dorsal and ventral streams: A framework for understanding aspects of the functional anatomy of language. *Cognition*, 92, 67–99.
- Hopfield, J.J. (1995). Pattern recognition computation using action potential timing for stimulus representation. *Nature*, 376, 33–36.
- Horwitz, B., Rumsey, J.M. & Donahue, B.C. (1998). Functional connectivity of the angular gyrus in normal reading and dyslexia. *Proceedings of the National Academy of Sciences*, 95, 8939–8944.
- Humphreys, G.W., Evett, L.J. & Quinlan, P.T. (1990). Orthographic processing in visual word identification. *Cognitive Psychology*, 22, 517–560.
- Iles, J., Walsh, V. & Richardson, A. (2000). Visual search performance in dyslexia. *Dyslexia*, 6, 163–177.
- Jordan, T.R., Patching, G.R. & Thomas, S.M. (2003). Assessing the role of hemispheric specialisation, serial-position processing, and retinal eccentricity in lateralised word recognition. *Cognitive Neuropsychology*, 20, 49–71.
- Kaufman, N.L. (1980). Review of research on reversal errors. *Perceptual Motor Skills*, 51, 55–79.
- Klingberg, T., Hedehus, M., Temple, E., Salz, T., Gabrieli, J.D., Moseley, M.E. & Poldrack, R.A. (2000). Microstructure of temporo-parietal white matter as a basis for reading ability: Evidence from diffusion tensor magnetic resonance imaging. *Neuron*, 25, 493–500.
- Lavidor, M. & Ellis, A.W. (2002). Orthographic neighborhood effects in the right but not in the left cerebral hemisphere. *Brain and Language*, 80, 63–76.
- Lavidor, M. & Walsh, V. (2004). Magnetic stimulation studies of foveal representation. *Brain and Language*, 88, 331–338.
- Leff, A. (2004). A historical review of the representation of the visual field in primary visual cortex with special reference to the neural mechanisms underlying macular sparing. *Brain and Language*, 88, 268–278.
- Legge, G.E., Mansfield, J.S. & Chung, S.T. (2001). Psychophysics of reading. XX. Linking letter recognition to reading speed in central and peripheral vision. *Vision Research*, 41, 725–743.
- Levelt, W.J. (2001). Spoken word production: A theory of lexical access. *Proceedings of the National Academy of Sciences*, 98, 13464–13471.
- Lieberman, I.Y., Shankweiler, D., Fischer, F.W. & Carter, B. (1974). Explicit syllable and phoneme segmentation in the young child. *Journal of Experimental Child Psychology*, 18, 201–212.
- Lisman, J.E. & Idiart, M.A. (1995). Storage of 7+/-2 short-term memories in oscillatory subcycles. *Science*, 267, 1512–1515.
- Livingstone, M.S., Rosen, G.D., Drislane, F.W. & Galaburda, A.M. (1991). Physiological and anatomical evidence for a magnocellular defect in developmental dyslexia. *Proceedings of the National Academy of Sciences*, 88, 7943–7947.
- Lorusso, M.L., Facoetti, A. & Molteni, M. (2004). Hemispheric, attentional, and processing speed factors in the treatment of developmental dyslexia. *Brain and Cognition*, 55, 341–348.

- Marslen-Wilson, W. & Warren, P. (1994). Levels of perceptual representation and process in lexical access: Words, phonemes, and features. *Psychological Review*, 101, 653–675.
- Mason, M. (1982). Recognition time for letters and non-letters: Effects of serial position, array size, and processing order. *Journal of Experimental Psychology*, 8, 724–738.
- Maunsell, J.H., Nealey, T.A. & DePriest, D.D. (1990). Magnocellular and parvocellular contributions to responses in the middle temporal visual area (MT) of the macaque monkey. *Journal of Neuroscience*, 10, 3323–3334.
- Mody, M., Studdert-Kennedy, M. & Brady, S. (1977). Speech perception deficits in poor readers: Auditory processing or phonological coding? *Journal of Experimental Child Psychology*, 64, 199–231.
- Nazir, T. (2003). On hemispheric specialization and visual field effects in the perception of print: A comment on Jordan, Patching, and Thomas. *Cognitive Neuropsychology*, 20, 73–80.
- New, B., Ferrand, L., Pallier, C. & Brysbaert, M. (forthcoming). Re-examining word length effects in visual word recognition: New evidence from the English Lexicon Project.
- Nice, D.S. & Harcum, E.R. (1976). Evidence from mutual masking for serial processing of tachistoscopic letter patterns. *Perception and Motor Skills*, 42, 991–1003.
- O'Regan, J.K., Levy-Schoen, A., Pynte, J. & Brugailiere, B. (1984). Convenient fixation location within isolated words of different length and structure. *Journal of Experimental Psychology: Human Perception and Performance*, 10, 250–257.
- Ostrosky-Solis, F., Efron, R. & Yund, E.W. (1991). Visual detectability gradients: Effect of illiteracy. *Brain and Cognition*, 17, 42–51.
- Paulesu, E., Frith, U., Snowling, M., Gallagher, A., Morton, J., Frackowiak, R.S. & Frith, C.D. (1996). Is developmental dyslexia a disconnection syndrome? Evidence from PET scanning. *Brain*, 119, 143–157.
- Peressotti, F. & Grainger, J. (1995). Letter-position coding in random constant arrays. *Perception and Psychophysics*, 57, 875–890.
- Peressotti, F. & Grainger, J. (1999). The role of letter identity and letter position in orthographic priming. *Perception and Psychophysics*, 61, 691–706.
- Pugh, K.R., Mencl, W.E., Shaywitz, B.A., Shaywitz, S.E., Fulbright, R.K., Constable, R.T., Skudlarski, P., Marchione, K.E., Jenner, A.R., Fletcher, J.M., Liberman, A.M., Shankweiler, D.P., Katz, L., Lacadie, C. & Gore, J.C. (2000). The angular gyrus in developmental dyslexia: Task-specific differences in functional connectivity within posterior cortex. *Psychological Science*, 11, 51–56.
- Roach, N.W. & Hogben, J.H. (2004). Attentional modulation of visual processing in adult dyslexia: A spatial-cuing deficit. *Psychological Science*, 15, 650–654.
- Schulte-Korne, G., Deimel, W., Bartling, J. & Remschmidt, H. (1998). Auditory processing and dyslexia: Evidence for a specific speech processing deficit. *Neuroreport*, 9, 337–340.
- Spencer, K. (2000). Is English a dyslexic language? *Dyslexia*, 6, 152–262.
- Stein, J. & Walsh, V. (1997). To see but not to read; the magnocellular theory of dyslexia. *Trends in Neuroscience*, 20, 147–152.
- Stuart, M. (1990). Processing strategies in a phoneme deletion task. *Quarterly Journal of Experimental Psychology A*, 42, 305–327.
- Tallal, P. (1980). Auditory temporal perception, phonics, and reading disabilities in children. *Brain and Language*, 9, 182–198.
- Tarkiainen, A., Cornelissen, P.L. & Salmelin, R. (2002). Dynamics of visual feature analysis and object-level processing in face versus letter-string perception. *Brain*, 125, 1125–1136.
- Tarkiainen, A., Helenius, P., Hansen, P.C., Cornelissen, P.L. & Salmelin, R. (1999). Dynamics of letter string perception in the human occipitotemporal cortex. *Brain*, 122, 2119–2132.
- Terepocki, M., Kruk, R.S. & Willows, D.M. (2002). The incidence and nature of letter orientation errors in reading disability. *Journal of Learning Disabilities*, 35, 214–233.
- Treiman, R. & Cassar, M. (1997). Can children and adults focus on sound as opposed to spelling in a phoneme counting task? *Developmental Psychology*, 33, 771–780.
- Vidyasagar, T.R. (2001). From attentional gating in macaque primary visual cortex to dyslexia in humans. *Progress in Brain Research*, 134, 297–312.
- Vidyasagar, T.R. (2004). Neural underpinnings of dyslexia as a disorder of visuo-spatial attention. *Clinical and Experimental Optometry*, 87, 4–10.
- Warren, P. & Marslen-Wilson, W.D. (1988). Cues to lexical choice: Discriminating place and voice. *Perception and Psychophysics*, 43, 21–30.
- Whitney, C. (2001). How the brain encodes the order of letters in a printed word: The SERIOL model and selective literature review. *Psychonomic Bulletin and Review*, 8, 221–243.

- Whitney, C. (2004a). Hemisphere-specific effects in word recognition do not require hemisphere-specific modes of access. *Brain and Language*, 88, 279–293.
- Whitney, C. (2004b). The Neural Basis of Structured Representations. PhD thesis, University of Maryland, College Park, MD.
- Whitney, C. (forthcoming). Supporting the serial in the SERIOL model.
- Whitney, C. & Berndt, R.S. (1999). A new model of letter string encoding: Simulating right neglect dyslexia. *Progress in Brain Research*, 121, 143–163.
- Whitney, C. & Lavidor, M. (2004). Why word length only matters in the left visual field. *Neuropsychologia*, 42, 1680–1688.
- Whitney, C. & Lavidor, M. (forthcoming). Facilitatory orthographic neighborhood effects: The SERIOL model account. *Cognitive Psychology*.
- Wolford, G. & Hollingsworth, S. (1974). Retinal location and string position as important variables in visual information processing. *Perception and Psychophysics*, 16, 437–442.

*Received 16 August 2004; revised version received 9 February 2005.*

**Address for correspondence:** Dr Carol Whitney, Department of Linguistics, University of Maryland, College Park, MD, 20742 USA. E-mail: [cwhitney@cs.umd.edu](mailto:cwhitney@cs.umd.edu)