

# Modeling the Effects of Formal Literacy Training on Language Mediated Visual Attention

**Alastair C. Smith (alastair.smith@mpi.nl)**

Max Planck Institute for Psycholinguistics, Wundtlaan 1,  
Nijmegen, 6525 XD, The Netherlands

**Padraic Monaghan (p.monaghan@lancaster.ac.uk)**

Department of Psychology, Lancaster University,  
Lancaster, LA1 4YF, U.K.

**Falk Huettig (falk.huettig@mpi.nl)**

Max Planck Institute for Psycholinguistics, Wundtlaan 1,  
Nijmegen, 6525 XD, The Netherlands

## Abstract

Recent empirical evidence suggests that language-mediated eye gaze is partly determined by level of formal literacy training. Huettig, Singh and Mishra (2011) showed that high-literate individuals' eye gaze was closely time locked to phonological overlap between a spoken target word and items presented in a visual display. In contrast, low-literate individuals' eye gaze was not related to phonological overlap, but was instead strongly influenced by semantic relationships between items. Our present study tests the hypothesis that this behavior is an emergent property of an increased ability to extract phonological structure from the speech signal, as in the case of high-literates, with low-literates more reliant on more coarse grained structure. This hypothesis was tested using a neural network model, that integrates linguistic information extracted from the speech signal with visual and semantic information within a central resource. We demonstrate that contrasts in fixation behavior similar to those observed between high and low literates emerge when models are trained on speech signals of contrasting granularity.

**Keywords:** The Visual World Paradigm, Connectionist Modeling, Visual Attention, Literacy.

## Introduction

Eye-tracking studies in which participants are presented simultaneously with spoken language and visual input (i.e. the visual world paradigm, Tanenhaus et al., 1995) have shown that information retrieved via both modalities is mapped at multiple levels of representation. Allopenna et al. (1998), for instance, presented participants with spoken words such as beaker and objects whose names contained word-initial or word-final overlapping phonological information (e.g., beetle, speaker) together with phonologically unrelated objects (e.g., carriage). They found that eye-movements were more likely to be directed to the phonologically related objects than to unrelated objects, indicating that during speech processing, phonologically related representations were co-activated and mapped onto phonological representations retrieved from viewing the co-present visual objects (see Huettig & McQueen, 2007, for further discussion). Related paradigms have demonstrated that semantic competitors are also co-activated during

listening to speech and attract increased overt attention (Yee & Sedivy, 2006; Huettig & Altmann, 2005)

These types of studies leave open one important question: What particular aspects of these representations affect participants' performance? Computational models have been proposed to reproduce the individual phonological and semantic effects on word processing. Allopenna et al. (1998), demonstrated that fixation probabilities during spoken word processing can be predicted by lexical activations in the TRACE model of spoken word recognition. Mayberry, Crocker and Knoeferle (2009) and Kukona and Tabor (2011) extended this work to predict fixation behavior during sentence processing from the integration of visual and linguistic information. Until recently, such models that simulate the interaction between visual and linguistic information did so with representations that were unable to capture fine-grained semantic, phonological or visual feature relationships and were therefore limited in their ability to examine effects of multimodal interactions in language processing. A recent model by Smith, Monaghan and Huettig (in press) based on the hub-and-spoke models of semantic processing which integrates visual, phonological and functional information within a central resource, replicated the intricate time course dynamics of eye fixation behavior reported in Huettig and McQueen (2007). The model highlights the role of differences in the computational properties of each modality's representational structure, demonstrating that such differences are sufficient to produce behavior consistent with multimodal effects reported in the Visual World Paradigm.

The question of how differential representational qualities of phonological and semantic properties affect word processing can also be approached by studying individual differences. Specifically studying participant populations that differ in the form of representation of each modality that they bring to the task. People with different levels of literacy are a critically important population in this regard. There is a well-established link between fidelity of phonological representations of words and development of

literacy (Hulme et al., 2012). Participants who are literate perform better at phonological segmentation or phoneme awareness tasks (Bowey, 2005), and there have been proposals both that literacy causes such improvements in phonological processing (Castles & Coltheart, 2004; Morais, Cary, Alegria, & Bertelson, 1979), as well as converse views that effective phonological processing results in improved reading (Muter, Hulme, Snowling, & Stevenson, 2004). An influential processing model in this literature is that experience of written forms of words results in a change in the granularity of the phonological processing of a word (Ziegler & Goswami, 2005), such that exposure to written words results in greater awareness of the individual phonemes of words, and without such exposure, listeners are more likely to process the sound of a word without a componential, phonological decoding.

In contrast, effects of literacy on semantic processing have been shown to be minimal and appear to be only quantitatively rather than qualitatively different (Da Silva et al., 2004; Reis & Castro-Caldas, 1997). Thus, literacy appears to affect lexical processing in a modality-specific manner.

In a recent study, Huettig, Singh and Mishra (2011) compared phonological and semantic competitor effects for Indian participants who had high and low levels of literacy due to poverty or other socioeconomic factors (but no known neurological or cognitive deficits), enabling a direct test of the extent to which the granularity of the phonological form of a word affects performance. In their study (Experiment 1), participants viewed a scene comprising objects representing a phonological onset competitor, a semantic competitor, and two unrelated distractors, and heard the target word spoken in a sentence context. They found that participants with low levels of literacy demonstrated no effects of phonological competitors, but substantial effects of semantic competitors when hearing words. In contrast, the participants with high levels of literacy were similar to the participants in a similar study with Dutch high literates (Huettig and McQueen, 2007) – demonstrating early looks towards objects named by phonological competitors and later looks toward semantic competitors.

We note that looks to the semantic competitors in the Huettig et al. (2011) study were reduced for the low literacy group, which is consistent with accounts of a general processing deficit (cf. Salthouse, 1996), and we return to this issue in the Discussion section.

We adapted our previous multi-modal model of fixation behavior in the visual world paradigm (Smith, Monaghan, & Huettig, in press) to test the explanatory adequacy of the hypothesis regarding granularity of phonological processing relating to different levels of literacy. We simulated the conditions of the experimental study by presenting visual object representations of phonological and semantic competitors, and two unrelated words and tracking the model's fixation of each of these objects as presentation of a target word unfolded. We adjusted the level of granularity

of the auditory presentation of the word to the model, predicting that a segmented phonological representation would result in early phonological competitor effects, but that less individuated phonological representations, consistent with accounts of phoneme awareness impairment in low-literacy groups, would result in reduced, or absent phonological effects. We also predicted that, consistent with the behavioural data, the later semantic competitor effects would be observed for the model regardless of the granularity of the auditory input to the model.

In order to isolate the effect of the granularity of auditory processing of the spoken word, we controlled for the overall similarity between words in terms of their auditory form, but varied whether the similarity was compositional and at the phoneme level within the model, whether it was sublexical but not at the phonological level, or whether it was not sublexical and represented only at the word level.

## Method

### Model

The models described in this paper are based on the model of language mediated eye-gaze presented in Smith, Monaghan and Huettig (in press). The general architecture of the model is shown in Figure 1.

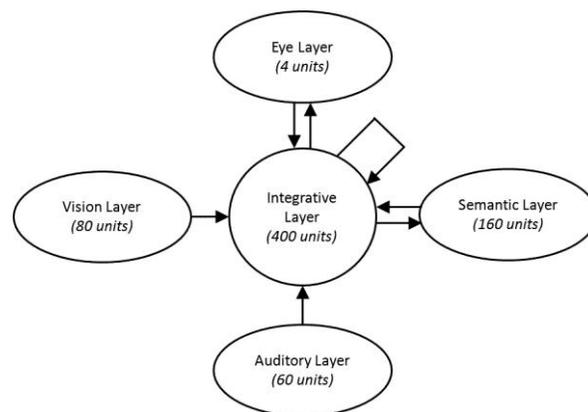


Figure 1: Network Architecture.

**Architecture** The network consists of four modality-specific layers which were fully connected to a central resource consisting of 400 units (see Figure 1). The model implements a hub-and-spokes model of multimodal integration, with input visual, auditory and semantic information about words, and output behavior of an “eye” layer which indicates the direction of the attentional focus of the model as a consequence of the combination of the modal inputs.

The vision layer (80 units) simulated the extraction of visual information from the surrounding environment, providing visual input to the system. It was divided into four slots, each defined by 20 processing units. Each slot corresponded to the visual information available at each of four possible locations within the visual field. The vision

layer was fully connected in a forward direction to the integrative layer.

Similarly the auditory layer provided input from the auditory modality, simulating the extraction of spoken information from the speech signal over time. The auditory layer was also fully connected to the central integrative layer in a forward direction.

The semantic layer consisted of 160 units, allowing the network to represent semantic features associated with a given object or spoken word. The semantic layer was fully connected to the integrative layer with activation flowing both from integrative units to semantic units and also back from semantic to integrative units.

The eye layer, to reflect the viewing behavior of the system, was also fully connected in both a forward and back direction to the central integrative layer. It consisted of four units, a unit for each location in the visual field represented in the vision layer. Activation of an eye unit was taken as representing the probability of fixating the location in the visual field associated with the given eye unit.

**Representations** An artificial corpus consisting of visual, auditory, and semantic representations for 200 items was constructed to train and test the network on multiple cross-modal tasks mapping between each of the modalities. A fundamentalist approach (Plaut, 2002) was taken in the construction of representations to ensure all aspects of the representations were controlled within the simulations.

Visual representations of named objects were implemented as 20 unit binary vectors, with each unit representing the presence or absence of a given visual feature for the object. Each object had approximately 10 units activated, which were selected at random, and balanced for their distribution across the set of all 200 items.

For the semantic representations, each item was represented in terms of 8 units active from a set of 160 semantic features, such that the overall set of semantic representations were fairly sparse, simulating semantically

distinct words. Semantically similar pairs of words each shared 4 of the 8 active units representing each item.

To simulate different grain-sizes of speech representation, three forms of auditory input were constructed, but with the overall similarity between representations controlled.

For the fine grained auditory processing, representing phonological segmentation of the spoken word by the listener, words were encoded as six phonemes, with phonemes implemented as sets of 10 units, from which five units were active. All words within the corpus were composed of phonemes taken from an inventory of 20 possible phonemes. To present the word an additional phoneme from the target word sequence was presented to the auditory layer at each time step.

To simulate sublexical representations of a coarser grain size (moderate), two 30 unit binary feature vectors were created for each word from which 15 units were active. Coarse grained representations were formed by 60 unit binary feature vectors of which 30 units were active.

Table 1: Mean cosine similarity of speech signal representations calculated between targets and distractors.

Grain Size	Distractor Type	Signal Overlap ( $\bar{x}, \sigma$ )		
		Onset	Rhyme	Word
Fine	Competitor	.18 (.07)	.50 (.13)	.34 (.07)
	Unrelated	.50 (.12)	.50 (.12)	.50 (.09)
Moderate	Competitor	.17 (.08)	.50 (.11)	.34 (.07)
	Unrelated	.51 (.10)	.51 (.10)	.50 (.07)
Coarse	Competitor	.34 (.10)	.34 (.10)	.34 (.07)
	Unrelated	.51 (.10)	.51 (.10)	.50 (.07)

Visual, semantic and auditory competitors were also embedded within the corpora for 40 target items. For visual competitors 10 of 20 visual features were shared with target items with  $p = 1$ , with the remaining features shared with  $p = 0.5$ . Semantic competitors shared 4 of 8 semantic features with target representations, while unrelated items shared a maximum of 1 semantic property with any other item.

Table 2: Temporal organization of events in training. Describes input and target representations provided in training trials.

Task	Visual Input		Auditory Input		Semantic Layer		Eye Layer	
	Activity	ts	Activity	ts	Activity	ts	Activity	ts
<b>Form to Semantics</b>	4 items selected at random from corpus	0 - 14	Time invariant noise provided as input	0 - 14	<i>Target: Target's Semantic representation</i>	3 - 14	Input: Only location of target active	0 - 14
<b>Speech to Semantics</b>	Time invariant noise provided as input	0 - 14	Phonology of target as staggered input	0 - 14	<i>Target: Target's Semantic representation</i>	5 - 14	No constraints on activation	-
<b>Speech to Location</b>	4 items selected at random from corpus	0 - 14	Phonology of target as staggered input	0 - 14	No constraints on activation	-	<i>Target: Only location of target active</i>	5 - 14
<b>Semantics to Location</b>	4 items selected at random from corpus	0 - 14	Time invariant noise provided as input	0 - 14	Input: Target's Semantic representation	0 - 14	<i>Target: Only location of target active</i>	2 - 14

Fine grained spoken word competitors were defined by an overlap in the initial two components of their speech signal. For the unrelated items, we ensured that this set of words did not share more than the first component of the word and that no items shared their initial nor final three components. For moderate grain size representations 2/3 of the initial 30 features of a competitor were shared with a target with  $p = 1$ , with remaining features overlapping with  $p = 0.5$ . Controls ensured all initial and final moderate grain vectors were unique. For coarse grain competitors 1/3 of all features were shared with the corresponding target with  $p = 1$ , with remaining features overlapping with  $p = 0.5$ . Defining competitors in this way lead to the contrasts in levels of similarity between representations across corpora as described in Table 1. Although the level of similarity between competitor-target and unrelated distractor-target is consistent across corpora at the word level, the distribution of overlap varies between implementations as a function of grain size.

**Model Training** The model was trained on four tasks (see table 2). Tasks were designed to simulate those performed by participants prior to testing through which associations between representations are acquired. The tasks were to map from visual representation to semantic representation, from auditory representation to the semantic representation, to activate the eye unit corresponding to the location of the item whose semantic representation is presented, and to activate the location of the item whose auditory representation is presented. Tasks were presented on a pseudo random basis with the task of mapping speech to location occurring four times less than other tasks. Items were selected from the corpus and assigned roles (target or distractor) and locations randomly. Initial connection weights were randomized and adjusted during training using recurrent back-propagation (learning rate = 0.05). Training was terminated after 850 000 trials.

## Results

In the following sections we report the performance of three categories of model 1) Fine, models trained and tested on representations that simulate extraction of fine grained structure within the speech signal; 2) Moderate, models trained and tested on representations that simulate extraction of moderate structure within the speech signal; 3) Coarse, models trained and tested on representations that simulate coarse grained structure within the speech signal. The following results represent performance averaged across five instantiations of each model. For each instantiation a new corpus was constructed on which it was then trained and tested each initialized with a different random seed.

### Pre-Test

Once trained all models were tested on their ability to complete each of the four training tasks for all items in the training corpus presented in all possible locations within the

visual field. All three categories of model displayed similar levels of performance across all four tasks. In mapping from speech to semantics, activation of the semantic layer was most similar (cosine similarity) to the target item for 100% of items for all models. When mapping from visual to semantic representations, activation in the semantic layer was most similar (cosine similarity) to that of the target for 98% of items in the case of coarse and fine grained models and 97% of items in the case of moderate models. When challenged to select the location of a target when presented with its corresponding auditory representation, the correct location was activated in both the coarse and fine models for 96% of items and 98% of items for moderate models. All models displayed equal performance when locating a target indicated by the presence of its semantic representation, selecting the correct location for 99% of items.

### Simulating Huettig, Singh and Mishra (2011)

The following conditions remained consistent across all simulations. Visual input was provided at time step (ts) 0 and remained until the end of each test trial (ts 29). We report the activation of each unit within the eye layer as a proportion of the total activation of all units within this layer. This proportion is taken to represent the probability of fixating  $p(\text{fix})$ , the associated location within the visual field. Word onset occurred at ts 5, with an additional component of the speech signal presented at each time step until the entire speech signal had unfolded (ts 10). Auditory input then remains fixed until the end of the test trial.

To simulate the conditions of Huettig, Singh and Mishra (2011) experiment 1, input to the models visual layer consisted of the visual representations of the target's auditory competitor and semantic competitor along with two unrelated distractors. The target word's auditory representation was presented as a staggered input to the auditory layer from ts 5. All models (fine, moderate and coarse) were tested on all 40 test sets embedded within the corpus (target, auditory competitor, semantic competitor and two unrelated distractors) in all 24 possible combinations of item and location. Figure 2 displays the change in  $p(\text{fixation})$  from ts 0 for each category of item (Aud = auditory competitor, Sem = semantic competitor, Control = unrelated distractor), averaged across all test trials.

For analysis ratios were calculated between the proportion of fixations to a given competitor and the sum of the proportion of fixations to both the competitor and distractors (see Huettig & McQueen, 2007). A value of 0.5 would indicate both items were fixated equally, a value greater than 0.5 would indicate increased fixation of the competitor and lower than 0.5 increased fixation of the distractor. Mean ratios were calculated across items and instantiations.

We conducted a 2-way ANOVA on the auditory competitor-distractor ratios with model as between-subject factor and time as within-subject factor for three theoretically-motivated time regions (preview, early and late). No significant differences were predicted during the

preview period which refers to the time between display onset (ts 0) until the first time step in which auditory information relating to the target word is able to influence output layers (ts 7). The remainder of test trials was divided equally into two time bins, an early (ts 8 - 18) and a late (ts 19 - 29) period as previous research had shown that auditory effects would occur (if at all) during the early but not the late period.

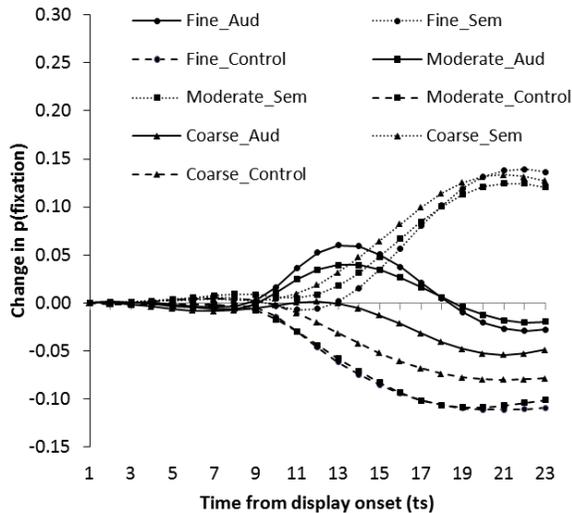


Figure 2: Change in fixation proportions for simulations of Huettig, Singh and Mishra (2011) Experiment 1.

There was a significant main effect of time,  $F(2, 234) = 38.155$ ,  $p < .001$ ,  $\eta^2 = .246$ , with auditory competitor-distractor ratios differing between preview and early time windows,  $F(1,238) = 39.387$ ,  $p < .001$ , and preview and late time windows,  $F(1,238) = 29.202$ , although there was no difference between early and late time windows. There was also a significant main effect of model,  $F(2, 117) = 4.467$ ,  $p = .014$ ,  $\eta^2 = .071$ , with the fine and medium models resulting in significantly more fixations to the phonological distractor than the coarse model, means = .544, .544, and .508, respectively. Critically, there was a significant interaction between model and time,  $F(4, 234) = 3.582$ ,  $p = .023$ ,  $\eta^2 = .058$ . The quadratic contrast effect for time was significant in the interaction,  $F(2, 117) = 5.074$ ,  $p = .008$ ,  $\eta^2 = .080$ , indicating that the models were more differentiated at the early time steps than during the preview or later time steps. Models did not differ significantly within the preview period. There was however a significant difference between fine and coarse models,  $F(1, 78) = 14.373$ ,  $p < .001$ , and coarse and moderate models,  $F(1, 78) = 9.544$ ,  $p = .003$ , in the early time window. The coarse model also differed from the fine  $F(1,78) = 4.286$ ,  $p = .042$ , and moderate model  $F(1,78) = 7.153$ ,  $p = 0.009$ , in the later time window. No difference was found between fine and moderate models in any time period.

A 2-way ANOVA was also conducted on semantic competitor-distractor ratios with model as between subject

factor and time as within-subject factor. Again we observed a main effect of time,  $F(2,234) = 230.642$ ,  $p < .001$ ,  $\eta^2 = .663$ , semantic competitor distractor ratios differed significantly between preview and early,  $F(1,238) = 59.607$ ,  $p < 0.001$  preview and late,  $F(1,238) = 243.403$ ,  $p < .001$  and early and late time windows,  $F(1,238) = 80.562$ ,  $p < .001$ . There was no main effect of model nor was there a significant interaction between model and time.

We then compared whether competitor-distractor ratios differed from chance (0.5) for each time step using one sample t-tests. The probability of fixating the auditory competitor first differed ( $p < 0.001$ ) from that of the distractor from time step 11 in both fine and moderate models and continued to differ for all subsequent time points. In contrast fixation of the auditory competitor by the coarse model only differed marginally ( $p < 0.1$ ) from the distractor item in time steps 13 – 17. Fixation of semantic competitors first differed significantly ( $p < 0.05$ ) from distractor levels at ts 12 and continued to differ for all remaining ts, this was the case for all models.

## Discussion

Our study aimed to examine the explanatory adequacy of the hypothesis that increased granularity of phonological processing, can account for the differences in fixation behavior between low and high literates observed in Huettig, Singh and Mishra (2011) Experiment 1. Our simulations demonstrate that increasing the grain size at which speech is processed can lead to a modulation of phonological effects. A model trained on representations of speech at the word level displayed only a marginal increase in fixation towards competitor items that overlapped in an auditory dimension, whereas models trained on componential, phoneme level representations or moderate grain size, sublexical components did display a significant increase in fixation of auditory competitors. Between model comparisons further demonstrated that the coarse grained implementation differed significantly from both fine and moderate grain models post word onset.

Interestingly, such comparisons did not display a graded effect of grain size, with fine and moderate models not differing in fixation proportions towards auditory competitors at any stage within test trials. There are two possible reasons for our failure to observe a graded effect. On the one hand, qualitative features of the data hint that given a larger corpus and hence test set such effects may be observable. One sample, left tailed t-tests comparing the ratio between the proportion of fixations towards auditory competitors in the moderate model and the sum of the proportion of fixations to the auditory competitor in the moderate and fine model indicate a significant difference at ts 13 – 16, ( $p < 0.05$ ), this difference can be observed in Figure 2.

On the other hand, it is conceivable that illiterates and low literates rely on very coarse grained structure within the speech signal. Although previous studies have shown that illiterates and low literates perform slightly better on

syllable awareness than on phonemic awareness tasks, they still tend to perform far worse than proficient readers. This may suggest that achieving even moderate granularity of phonological processing may not be rapid. The results of our simulations could be interpreted as reflecting that when a moderate grain size of phonological processing is achieved performance improves rapidly and becomes similar to fine-grained models.

Our results also demonstrate that increased granularity does not necessarily lead to a decrease in semantic effects as observed in Huettig, Singh and Mishra (2011). Although our simulations indicate phonological effects could be modulated by an increase in the grain size, an additional mechanism is needed to create the distinction between semantic effects observed across populations. A reduction in general processing speed in the illiterate population has been offered to account for differences in performance on a large variety of cognitive tasks (Salthouse, 1996). This potentially offers an explanation for a reduction in both auditory and semantic competitor effects. A general processing deficit for low literates, could be implemented by adding noise across semantic representations, representing a reduction in the fidelity of such representations. Adding noise in this manner would result in a general reduction of semantic competitor effects, however it is less clear whether the introduction of noise could also lead to the elimination rather than a general reduction of the phonological effect as observed in illiterate performance. As the authors acknowledge, behavior observed in Huettig et al (2011) suggests that the qualitative changes to the phonological competitor effects and the semantic competitor effects are distinct. Teasing apart the factors underlying observed differences in behaviour between populations is far from trivial, however explicit implementations such as the one described in this paper provide a means of testing the plausibility of proposed explanations.

## References

- Allopenna, P. D., Magnuson, J. S., and Tanenhaus, M. K. (1998). Tracking the time course of spoken word recognition using eye movements: evidence for continuous mapping models. *Journal of Memory and Language*, 38, 419–439.
- Bowey, J. A. (2005). Predicting individual differences in learning to read. In M. J. Snowling & C. Hulme (Ed.), *The science of reading: A handbook* (pp. 153–172). Oxford, England: Blackwell.
- Castles, A., & Coltheart, M. (2004). Is there a causal link from phonological awareness to success in learning to read? *Cognition*, 91, 77–111.
- Da Silva, C., Petersson, K. M., Faísca, L., Ingvar, M., and Reis, A. (2004). The effects of literacy and education on the quantitative and qualitative aspects of semantic verbal fluency. *Journal of Clinical Experimental Neuropsychology*, 26, 266–277.
- Huettig, F., & Altmann, G. T. (2005). Word meaning and the control of eye fixation: Semantic competitor effects and the visual world paradigm. *Cognition*, 96(1), B23–B32.
- Huettig, F., and McQueen, J. M. (2007). The tug of war between phonological, semantic, and shape information in language-mediated visual search. *Journal of Memory and Language*, 54, 460–482.
- Huettig, F., Singh, N., & Mishra, R. K. (2011). Language-mediated visual orienting behavior in low and high literates. *Frontiers in Psychology*, 2, 285.
- Hulme, C., Bowyer-Crane, C., Carroll, J.M., Duff, F.J., & Snowling, M.J. (2012). The causal role of phoneme awareness and letter-sound knowledge in learning to read: Combining intervention studies with mediation analyses. *Psychological Science*, 23, 572–577.
- Kukona, A., & Tabor, W. (2011). Impulse processing: A dynamical systems model of incremental eye movements in the visual world paradigm. *Cognitive Science*, 35, 1009–1051.
- Mayberry, M., Crocker, M. W., & Knoeferle, P. (2009). Learning to attend: A connectionist model of the coordinated interplay of utterance, visual context, and world knowledge. *Cognitive Science*, 33, 449–496.
- Morais, J., Cary, L., Alegria, J., & Bertelson, P. (1979). Does awareness of speech as a sequence of phones arise spontaneously? *Cognition*, 7, 323–331.
- Muter, V., Hulme, C., Snowling, M. J., & Stevenson, J. (2004). Phonemes, rimes, vocabulary and grammatical skills as foundations of early reading development: Evidence from a longitudinal study. *Developmental Psychology*, 40, 665–681.
- Plaut, D. C. (2002). Graded modality-specific specialization in semantics: A computational account of optic aphasia. *Cognitive Neuropsychology*, 19, 603–639.
- Reis, A., and Castro-Caldas, A. (1997). Illiteracy: a bias for cognitive development. *Journal of the International Neuropsychological Society*, 3, 444–450.
- Salthouse, T. A. (1996). The processing speed theory of adult age differences in cognition. *Psychological Review*, 103, 403–428.
- Smith, A., Monaghan, P., & Huettig, F. (in press). Multimodal interaction in a model of visual world phenomena. In *Computational Models of Cognitive Processes: Proceedings of the 13th Neural Computation and Psychology Workshop*. World Scientific Publishing Company.
- Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science*, 268, 1632–1634.
- Yee, E., & Sedivy, J. C. (2006). Eye movements to pictures reveal transient semantic activation during spoken word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 32, 1–14.
- Ziegler, J. C., & Goswami, U. (2005). Reading acquisition, developmental dyslexia, and skilled reading across languages: A psycholinguistic grain size theory. *Psychological Bulletin*, 131, 3–29.