Cognitive Psychology 75 (2014) 28–54

Contents lists available at ScienceDirect

Cognitive Psychology

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

Literacy effects on language and vision: Emergent effects from an amodal shared resource (ASR) computational model

Alastair C. Smith [a,b,*], Padraic Monaghan [c], Falk Huettig [a,d]

[a] Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands
[b] International Max Planck Research School for Language Sciences, Nijmegen, The Netherlands
[c] Department of Psychology, Lancaster University, Lancaster, UK
[d] Donders Institute for Brain, Cognition, and Behaviour, Radboud University, Nijmegen, The Netherlands

A R T I C L E I N F O

Article history:
Accepted 24 July 2014

Keywords:
Literacy
Computational modelling
Visual attention
Speech processing
Eye movements
Visual world paradigm

A B S T R A C T

Learning to read and write requires an individual to connect additional orthographic representations to pre-existing mappings between phonological and semantic representations of words. Past empirical results suggest that the process of learning to read and write (at least in alphabetic languages) elicits changes in the language processing system, by either increasing the cognitive efficiency of mapping between representations associated with a word, or by changing the granularity of phonological processing of spoken language, or through a combination of both. Behavioural effects of literacy have typically been assessed in offline explicit tasks that have addressed only phonological processing. However, a recent eye tracking study compared high and low literate participants on effects of phonology and semantics in processing measured implicitly using eye movements. High literates' eye movements were more affected by phonological overlap in online speech than low literates, with only subtle differences observed in semantics. We determined whether these effects were due to cognitive efficiency and/or granularity of speech processing in a multimodal model of speech processing – the amodal shared resource model (ASR, Smith, Monaghan, & Huettig, 2013a,b). We found that cognitive efficiency in the model had only a marginal effect on semantic processing and did not affect performance for phonological processing, whereas fine-grained versus coarse-grained

* Corresponding author. Address: Max Planck Institute for Psycholinguistics, P.O. Box 310, 6500 AH Nijmegen, The Netherlands.

E-mail address: alastair.smith@mpi.nl (A.C. Smith).

http://dx.doi.org/10.1016/j.cogpsych.2014.07.002

phonological representations in the model simulated the high/low literacy effects on phonological processing, suggesting that literacy has a focused effect in changing the grain-size of phonological mappings.

1. Introduction

Approximately 16% of the world's adult population are illiterate, defined as "the ability to read and write with understanding a simple statement related to one's daily life" (UNESCO Institute for Statistics., 2013). Learning to read has a profound effect on cognitive processing, resulting in qualitative changes to the representation of phonological information about words, but also correlating with a general increase in cognitive processing performance. Much of our understanding of language processing is based on data and theoretical and computational models only of literate participants, but a full understanding of language comprehension and production must also take into account the role of literacy in processing. Previous models of literacy effects on language processing have not effectively distinguished between accounts based on a general cognitive ability increase and more specific phonological processing changes.

Here, we test an implemented computational model of language processing that was previously applied only to data from literate participants. We extended the model to simulate both the general cognitive processing account as well as the phonological representation account in order to account for data from literate and illiterate participants in language processing tasks. We first review the two theoretical accounts of effects of literacy on language processing – the phonological processing change and the general cognitive processing accounts – before describing previous models of effects of literacy on language processing. We then present the advantages of a language processing task that tests online, implicit processing of information between vision, phonology and semantics in order to examine the effects of literacy on the language processing system, before presenting our model's design and results.

1.1. Changes to phonological representations and literacy

The aspect of speech processing for which there has been most exploration for an influence of literacy is in the domain of phonological awareness, defined as "one's degree of sensitivity to the sound structure of oral language" (Anthony & Francis, 2005). There is substantial evidence indicating that, over the course of development, individuals become increasingly sensitive to smaller linguistic units within the speech signal. Children first gain awareness of larger units such as syllables before they are able to display an awareness of smaller units such as onsets and rhymes (Alcock, Ngorosho, Deus, & Jukes, 2010; Anthony & Francis, 2005; Goswami, 2003). However, debate remains as to the cause of this improvement. Firstly, what is the role of literacy acquisition? Is perceptual categorisation of speech sounds dependent on reading acquisition (Burnham, 2003)? Does literacy lead to a finer tuning of perceptual categories and, consequently, improvements in the precision of phoneme identification (Hoonhorst et al., 2011; Serniclaes, Ventura, Morais, & Kolinsky, 2005)? Or does literacy not play a crucial role, instead is it that the fidelity of phonological representations increases across development driven by the need to differentiate, within an increasingly large lexicon, between an increasing number of phonologically similar items (Garlock, Walley, & Metsala, 2001; Storkel, 2002)?

There is growing evidence that for (at least) explicit awareness of fine grain phonological units, individuals require exposure to alphabetic literacy training. Experiments that require children to make explicit judgements regarding a word's phonological structure show that children perform largely at chance prior to literacy training, however once engaged in training their performance on such tasks greatly improves (Alcock et al., 2010; De Jong & Van Der Leij, 2003; Hulme, Snowling, Caravolas, &

Carroll, 2005; Morrison, Smith, & Dow-Ehrensberger, 1995; Treiman & Zukowski, 1991). Critically, similar tests have been conducted on illiterate adults with such individuals also displaying chance level performance on tasks requiring explicit phoneme manipulation or judgments (Adrián, Alegria, & Morais, 1995; Loureiro, Willadino Braga, Souza, Queiroz, & Dellatolas, 2004; Morais, Cary, Alegria, & Bertelson, 1979; Scliar-Cabral, Morais, Nepomuceno, & Kolinsky, 1997). It has also been observed that late literates (individuals who learn to read in adulthood) although displaying improved performance compared to illiterates on phonological awareness tasks still perform worse than early literates (individuals who learn to read during childhood) (Morais, Bertelson, Cary, & Alegria, 1986). Although performance of illiterates on phonological awareness tasks has been shown to be very poor, illiterates display improved performance (although performance is still lower than literates) on metaphonemic judgement tasks (Syllable Detection: Morais, Content, Cary, Mehler, & Segui, 1989; Rhyme Awareness: Adrián et al., 1995; Morais et al., 1986; Phonological Length: Kolinsky, Cary, & Morais, 1987). Such data indicates that increases in phonological awareness, displayed by literate children, do not emerge simply as a result of greater exposure to spoken language or as the system matures, instead this evidence implicates literacy training as the critical factor in enabling explicit phonological awareness.

What is less clear is the impact of literacy on online speech processing. The above studies require participants to make explicit judgements regarding phonological properties of words. Based on this evidence alone it is not possible to say whether the progression towards explicit knowledge of more fine-grained components in the speech signal is also mirrored in an individual's implicit abilities when processing speech online.

Evidence for effects of literacy for online speech processing is less prevalent and less conclusive. Reis and Castro-Caldas (1997) observed that illiterates performed worse than literates on a pseudo word repetition task, whereas both populations performed equally well when repeating real words, suggesting that sub-lexical representations of spoken words were less readily accessible to the illiterate participants. Literacy has also been shown to influence categorical perception in speech. Serniclaes et al. (2005) showed that literates displayed sharper boundary precision in response to *ba-da* contrasts than illiterates, an effect that correlated with reading level (Hoonhorst et al., 2011). Such findings are consistent with an increase in the fidelity of phonological representation as a consequence of literacy, yet could instead indicate a more subtle refinement of categorical boundaries rather than confirming a prior absence of phoneme level representations (Burnham, 2003).

Although this evidence is largely consistent with literacy leading to more fine grained processing of the speech signal, it provides little insight regarding the stages in online speech processing affected by literacy training, for example does literacy lead to changes in early perceptual processing or are observed differences dependent on feedback from later activated orthographic knowledge? Such insight is important as phonological processing occurs rapidly with effects often transitory and dynamic in nature, so understanding the timing of these effects may provide the necessary evidence required to isolate differences in underlying cognitive processing.

Behavioural evidence is scarce regarding time-course effects of literacy, though one study that isolates timing differences (Ventura, Kolinsky, Querido, Fernandes, & Morais, 2007) compared performance of literates and illiterates on a picture word interference task in which named pictures shared only the first phoneme with a spoken word. Results showed a phonological priming effect for both groups. However, illiterates only displayed an effect at later SOAs. This is compatible with more coarse grained processing of the speech input in illiterates, as it could be argued that more of the speech signal needs to unfold before overlapping representations are activated and can exert an influence on behaviour. ERP (Event-related potential) data has also provided a productive means of probing time-course effects of literacy on online speech processing. Such studies demonstrate an early influence of orthography during spoken word processing, critically with effects observed in windows prior to points that are classically viewed as the time point of lexical access (Semantic categorisation task: Pattamadilok, Perre, Dufau, & Ziegler, 2009; Lexical decision task: Perre, Midgley, & Ziegler, 2009a; Perre, Pattamadilok, Montant, & Ziegler, 2009b; Perre & Ziegler, 2008).

Ziegler and Ferrand (1998) suggest that the mechanisms underlying the effects of orthography on online speech processing are that following literacy training orthographic representations are activated online when processing spoken words and it is such online activation that leads to effects of orthography on speech processing. Neuroimaging evidence consistent with this hypothesis can be found in Dehaene et al. (2010) in which online speech processing tasks in literates, but not illiterates, were observed to activate brain regions associated with orthographic processing. However, such evidence is not incompatible with an alternative restructuring hypothesis in which the process of learning orthographic mappings leads to adaptation in other language processing regions. For example, phonological processing regions may be restructured so that processing reflects characteristics of orthographic representations, such as being finer grained (Muneaux & Ziegler, 2004; Taft, 2006; Taft & Hambly, 1985). Neural data has also provided evidence in support of a restructuring account, by isolating effects of literacy on speech processing to regions associated with phonological processing. For example, Perre et al. (2009b) localized the source of orthographic consistency effects during spoken word recognition observed in ERP data to classic phonological processing regions (left BA40). Further, Pattamadilok, Knierim, Duncan, and Devlin (2010) demonstrated that orthographic consistency effects during auditory lexical decision tasks can be removed when disturbing processing in these phonological processing regions (left supramarginal gyrus) using repetitive transcranial magnetic stimulation, while they were not affected by disturbance of orthographic processing regions (left ventral occipito-temporal cortex).

Psycholinguistic grain size theory (Ziegler & Goswami, 2005) offers a processing level model that connects exposure to the written forms of words to increased granularity of phonological processing. It is also largely consistent with the behavioural and neural data presented earlier. Grain size theory proposes that learning to map between orthographic and phonological representations leads to a restructuring of phonological representations and is necessary to develop awareness of fine grained structure in the phonological lexicon, with the nature of the correspondence between orthographic units and phonological units within a given language determining the granularity of restructuring for that language. Ziegler et al. (2010) found a relationship between phonological awareness and reading performance across a range of alphabetic orthographies in children in second grade of school. Though this relation was found to be stronger for more opaque orthographies, this may be because readers of transparent orthographies develop ceiling effects in phonological awareness skills earlier in reading exposure than readers of opaque orthographies (Caravolas, Volin, & Hulme, 2005). Nonetheless, training on orthographies, where the correspondence between individual phonemes and letters is largely consistent, as in the case of alphabetic languages, is likely to lead to finer-grained phonological representations, in comparison to orthographies where orthographic units correspond only to larger, coarser-grained phonological units comprising multiple phonemes, for example in logographic languages. Awareness of larger units within words (i.e. syllables, onsets, rhymes) may proceed without literacy training; however for awareness of fine grain units to emerge (i.e. phonemes) it has been proposed that explicit training is necessary. Evidence in support of this position comes from observed similarities in processing between illiterates and logographic literates, for example Chinese literates, where there is little systematic correspondence between orthographic representations and the sequence of speech sounds that constitute their spoken form (Brennan, Cao, Pedroarena-Leal, McNorgan, & Booth, 2013; Cao et al., 2011; Cheung, Chen, Lai, Wong, & Hills, 2001; Ho & Bryant, 1997; Huang & Hanley, 1995, 1997; McBride-Chang, Bialystok, Chong, & Li, 2004; Read, Yun-Fei, Hong-Yin, & Bao-Qing, 1986; Shu, Peng, & McBride-Chang, 2008).

In our model of online speech processing we test a phonological restructuring hypothesis consistent with psycholinguistic grain size theory, in which learning to map between orthographic and phonological representations leads to changes in the granularity of phonological processing that reflect the structure of the orthographic system on which the system is trained. Therefore, training on alphabetic languages, in which there is a regular mapping between individual orthographic and phonological units leads to more fine grained phonological processing.

1.2. Cognitive efficiency and literacy

The effects of literacy, however, have not been isolated only to the domain of phonological processing. Historically, illiteracy has been linked to reduced performance on a range of cognitive tasks, e.g., visual perception (Luria, 1976), reasoning (Levi-Bruhl, 1923), and memory (Vygotsky, 1978). However, isolating the role of literacy from other factors such as pre-existing cognitive deficits or increased exposure to formal schooling is a substantial challenge. Yet, more recent studies that have attempted to control for such factors have continued to demonstrate a link between literacy and changes in cognitive performance on tasks that extend beyond the domain of phonological processing.

Performance on standardized memory tasks has been observed to differ between literate and illiterate groups. Specifically, illiterates display worse performance than literates on digit span tasks, in which participants are required to repeat a sequence of digits (Reis, Guerreiro, & Petersson, 2003). In the domain of semantic processing, performance on semantic fluency tasks, in which participants are required to produce as many items from a pre-specified semantic category as possible, has also been shown to differ between literate and illiterate groups (Kosmidis, Tsapkini, Folia, Vlahou, & Kiosseoglou, 2004; Reis & Castro-Caldas, 1997), though see da Silva, Petersson, Faísca, Ingvar, and Reis (2004) for an alternative account of semantic effects being due to differences in general knowledge.

Effects on visual processing have also been observed and appear to extend as far as low level perceptual processing. For example in a recent study by Szwed, Ventura, Querido, Cohen, and Dehaene (2012), illiterates performed worse than literates on a contour integration task, in which participants were required to indicate the direction of an image when the image was distorted by low level visual noise. In a visual target detection task in which participants were required to touch red squares placed among yellow squares on a computer screen, illiterates were shown to be slower and less accurate than literates (Bramao et al., 2007). However, such effects do not seem to be driven purely by low level perceptual differences. A more recent study examining visual search behaviour in literate and illiterate groups also observed slower performance in illiterate groups (Olivers, Huettig, Singh, & Mishra, 2014), yet demonstrated that the observed difference in behaviour was largely accounted for by low literates needing more time between fixating the target and producing a required motor response. A possible explanation for this consistent reduction in performance displayed by illiterates across many cognitive domains would be that literacy leads to a general increase in efficiency of cognitive processing.

General processing speed (Kail & Salthouse, 1994; Salthouse, 1996) has been shown to correlate with performance on a wide range of cognitive tasks (Kail & Salthouse, 1994; Li et al., 2004; Salthouse, 2005), and has been proposed to be the mechanism of increased cognitive efficiency as a consequence of literacy training. For instance, Stoodley and Stein (2006) showed that literacy skills correlated with a general increase in speed of performance on a pure motor task. It has been suggested that general processing speed is related to the rate at which information propagates from one node in a network to another (Kail & Salthouse, 1994; Salthouse, 1988). Such arguments are consistent with recent research in the field of neuroscience, into the effects of myelination in the human brain. Measures of myelination and white matter integrity have been shown to be reflected in the efficiency (Deary et al., 2006; Engel, Fries, & Singer, 2001; Li et al., 2009) and the speed (Gutiérrez, Boison, Heinemann, & Stoffel, 1995; Madden, Bennett, & Song, 2009; Penke et al., 2010; Tolhurst & Lewis, 1992; Waxman, 1980) of information processing, with such factors shown to modulate performance on a range of cognitive tasks (Deary et al., 2006; Li et al., 2009; Turken et al., 2008). Critically, myelination has been shown to be modifiable by experience (see Fields, 2008) and to increase as a result of learning (Bengtsson et al., 2005) and therefore has the potential for modulation by environmental variables such as exposure to literacy training. Studies have indeed shown that myelination of brain regions associated with language processing coincides with vocabulary acquisition (Pujol et al., 2006).

1.3. Models of literacy effects on language processing

Many of the most influential cognitive models of speech processing do not implement a role for orthographic knowledge (e.g., Cohort Model, Marslen-Wilson & Tyler, 1980; MERGE, Norris, McQueen, & Cutler, 2000; Shortlist B, Norris & McQueen, 2008; TRACE, McClelland & Elman, 1986), however there is one model of which we are aware that provides insight into potential effects of literacy on phonological processing of words. Harm and Seidenberg (1999) compared behaviour of a computational model trained to generate stable phonological representations of monosyllabic words, where the phonological representation was input as a set of phoneme features representing each phoneme in the word, to a model that in addition mapped orthographic representations onto the phonological representations, and assessed the effect of this literacy on the model's performance on a range of single-word phonological processing tasks. They found that when one phoneme within a word was affected by noise, the literate model was better able to restore the phoneme. They also found that the literate model represented words with the same rhymes as more similar in terms of the internal state of the model than the illiterate model.

These simulations provide an explicit description of how learning orthographic mappings can lead to emergent effects on phonological processing, modulating the componential nature of processing. They observed that changes to connection weights within the phonological network as a consequence of literacy training were greater for within-segment weights (within rhyme and within onset connections), rather than intersegmental weights (those crossing the onset-rhyme boundary). This they argue is consistent with evidence indicating that literacy leads to increased sensitivity to smaller phonological units. Harm and Seidenberg's (1999) model suggests that increased componentiality may be a consequence of literacy, consistent with the psycholinguistic grain-size theory and the restructuring hypothesis. However, the model represented phonology in terms of individual phonemes, thus increasing the chances that the model will discover phoneme-level representations. The model also did not test the potential effect of general cognitive processing advantages as a consequence of literacy within the model. For instance, similar observations of increased componentiality within the model could equally be a consequence of a model with greater fidelity of representations rather than (due only to) changes in the granularity of processing. For example, a model that possesses noisier representations is likely to perform worse on restoration tasks and may also represent words with the same rhymes as less similar.

1.4. Explicit and implicit phonological processing tasks

The above behavioural and computational studies provide substantial converging evidence for a connection between acquisition of literacy and the fidelity of phonological representations of words (Dijkstra, Roelofs, & Fieuws, 1995; Chéreau, Gaskell, & Dumay, 2007; Hulme, Bowyer-Crane, Carroll, Duff, & Snowling, 2012; Kolinsky, Pattamadilok, & Morais, 2012; Ventura, Morais, Pattamadilok, & Kolinsky, 2004; Ziegler & Ferrand, 1998). However, these previous studies have generally focused on explicit tasks addressing manipulations of the phonological forms of isolated words, and there is little extant evidence for behavioural consequences that may result from differences in phonological activation during online speech processing. This is important ecologically because these previous studies have focused either on manipulations of the phonological representation itself, or on the extent to which phonological representations are similar to one another, rather than the use that the language processing system makes of these representations. Language processing involves combinations of phonological, orthographic, and semantic representations in interaction with sensory input about the environment, and so determining the effects of literacy on language processing should take this complexity into account, rather than only focusing on one small aspect of the language processing system. Implementing the complexity of the system also permits testing the various accounts of effects of literacy on representations other than only phonological forms, which may prove important for distinguishing competing accounts based on cognitive efficiency or grain-size of phonological representations.

One way in which use of phonological representations can be studied is through the visual world paradigm (Cooper, 1974; Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995). In studies of this kind, participants are presented with a visual display while simultaneously hearing a spoken utterance, and as these events unfold their eye gaze is recorded. The paradigm has been previously used to examine integration of information between visual and linguistic representations by manipulating the relationships between items in the visual scene and words presented in the auditory stimulus (for review, see Huettig, Rommers, & Meyer, 2011).

Huettig, Singh, and Mishra (2011) conducted a visual world paradigm study with two populations in India both of which were native Hindi speakers (all materials were in Hindi). One was a high literate population, comprising undergraduate university students, and the other was a low literate population, that complied with the UNESCO definition of illiterate (provided earlier in this paper) yet who were fully integrated within Indian society. Low literates were employed and displayed no obvious social, cognitive or neurological deficits. The critical difference between populations was the amount of exposure to formal education. In their Experiment 1, participants listened to sentences containing a target word and were shown scenes containing a semantic competitor (item that shares a subset of its semantic features with the spoken target word; e.g. target = beaker, semantic competitor = fork), a phonological onset competitor (item that shares its phonological onset with the spoken target word; e.g. target = beaker, phonological competitor = beetle) and two unrelated distractors (items that did not share either a phonological or semantic relationship with the spoken target word). In their Experiment 2, scenes only displayed a phonological competitor and three unrelated distractors. On experimental trials visual scenes did not contain the named item, whereas filler trials contained scenes displaying pictures of the spoken target word in addition to three unrelated distractors. In both experiments participants performed a look and listen task, this simply required participants to look at the scenes while listening to the spoken sentence with no additional explicit task.

The results of Huettig et al.'s (2011) two experiments are displayed in Fig. 1. When presented with scenes containing phonological onset competitors and semantic competitors, high literates looked first at phonological competitors and then later at semantic competitors once information within the unfolding speech mismatched with the name of the phonological competitor, replicating earlier research (Huettig & McQueen, 2007). Low literates on the other hand only displayed increased fixation of semantic competitors, at no point fixating phonological competitors consistently more than unrelated distractors. Also, overall fixation of semantic competitors by low literates was lower than that

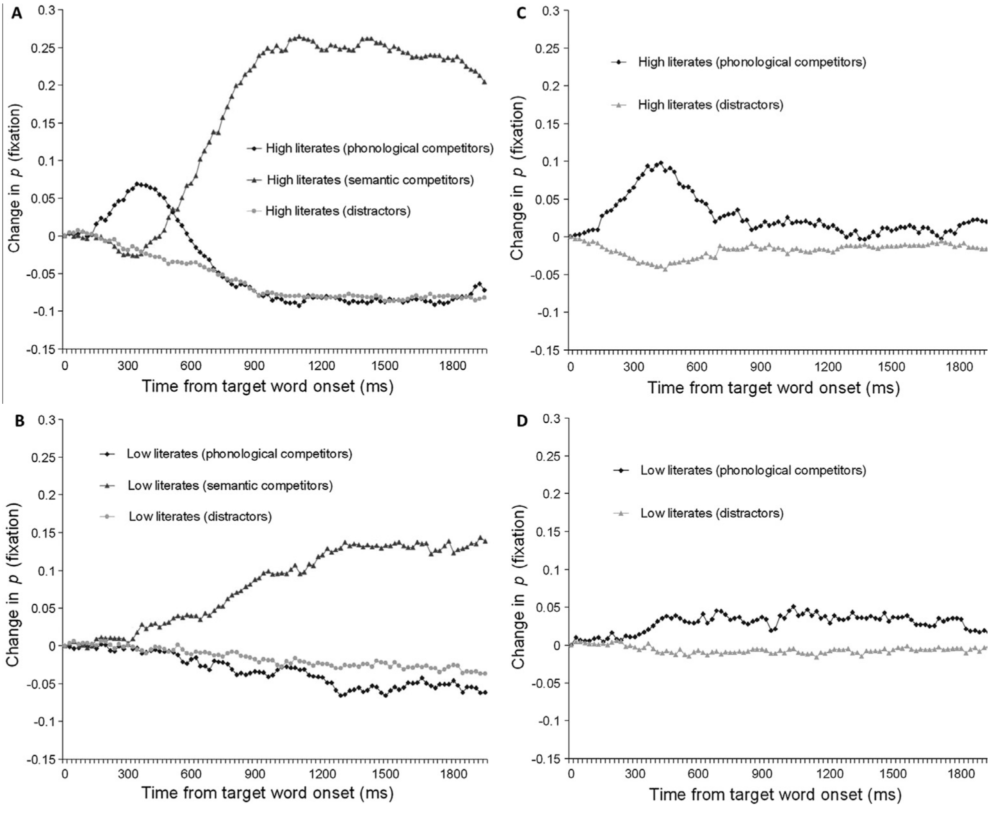

Fig. 1. Results of Huettig et al. (2011). Charts display the change in fixation proportions for high (A and C) and low (B and D) literates when presented with scenes containing either a phonological competitor, semantic competitor and unrelated distractors (A and B), or a phonological competitor and unrelated distractors (C and D). (Figures as published in Huettig et al. (2011). Language-mediated visual orienting behaviour in low and high literates. Frontiers in Psychology, 2, 285. doi: http://dx.doi.org/10.3389/fpsyg.2011.00285).

displayed by high literates. In the second experiment, when participants viewed scenes containing only a phonological onset competitor accompanied by unrelated distractors, high literates again displayed a pattern of fixation toward the phonological competitor that tightly mirrored the phonological overlap in the speech signal. Low literates on the other hand, unlike in Experiment 1, did display increased fixation of the phonological competitor compared to unrelated distractors, but in contrast to high literates, their fixations of phonological competitors were not tightly time locked to the unfolding speech signal. Low literates looked marginally more at the phonological competitor over the first 1000 ms post word onset but did not display the rapid increase and decrease in looks towards this category of item in response to signal overlap as shown by high literates.

These results support a qualitative difference between high and low literate populations in their use of phonological information, whereas only a small quantitative difference in terms of semantic processing. In terms of the two theories of effects of literacy on language processing – the fine-grained phonological representation and the cognitive efficiency theories – the results may be consistent with either. For high literate participants, they are sensitive to the phonological overlap between representations of words at the point at which the phonology of competing visual objects applies. For low literate participants, such sensitivity is not observed suggesting that fine-grained phonological distinctions between words are a consequence of literacy. The low literates may be instead biased toward mapping at a semantic level and only map at a phonological level when pushed by restrictions on the nature of the representational overlap in the environment. Huettig et al. (2011) conjecture that literacy results in these effects by strengthening existing phonological representations and providing a tighter coupling between phonological representations activated by the visual environment and phonological representations activated by spoken language input.

However, an alternative perspective is that the results are a consequence of greater efficiency in cognitive processing. Turken et al. (2008) highlight the efficiency of signal transmission across white matter tracts as a particularly significant factor in determining performance on tasks that require the complex integration of information from multiple operations. Therefore, the processes driving language mediated eye gaze, in which information from auditory, visual, semantic and eye gaze processing regions must be tightly integrated and used to coordinate behaviour, may be greatly influenced by such a variable. In terms of this theory, the observed results are then just due to the greater effectiveness and fidelity of the phonological representations for the high literate participants, rather than a qualitative difference in the grain-size of the processing.

The results of Huettig et al. (2011) alone do not provide a means of testing these two alternative hypotheses for the effects of literacy training on online language processing. However, the current study aims to demonstrate that through combining the rich behavioural data provided by the visual world paradigm with a computational model of language meditated visual attention it is possible to gain traction on such questions of effects of literacy on language processing that to date have eluded researchers.

The visual world paradigm is well suited to the investigation of phonological processing effects as it enables researchers to test effects not only of phonological processing but also of semantic effects on processing. Capturing these two effects in a single online task is required in order to distinguish between these alternative theories. Further, the use of a dense sampling method such as eye gaze, to compare between literate and illiterate behaviour, allows examination of moment by moment performance providing additional constraints on underlying processing differences during online speech processing that would otherwise be lost in more global measures. Within the current modelling approach, we use this rich behavioural measure to compare alternative theoretical explanations for the observed data by testing the behaviour of explicit implementations of each theory in a computational model.

In order to test these theories of literacy effects on language processing, we adapted the amodal shared resource (ASR) model of language mediated visual attention (Smith, Monaghan, & Huettig, 2013a,b). This model offers an explicit description of the information and processes that drive complex multimodal behaviour in language processing. The model has previously been shown to replicate a broad range of word level effects, displayed by literate populations, reported in the visual world literature (see Table 1). For example, the model successfully replicates contrasts across modalities in the effect of representational overlap on fixation behaviour of literate populations. The model replicates

Table 1
Table presenting visual world data successfully replicated by the ASR model of language mediated visual attention (Smith et al., 2013a,b).

Study	Scene			
Authors (year)	Item 1	Item 2	Item 3	Item 4
Allopenna et al. (1998)	Target	**Onset**	**Rhyme**	Distractor
Dahan and Tanenhaus (2005)	Target	**Visual**	Distractor	Distractor
Huettig and Altmann (2007)	**Visual**	Distractor	Distractor	Distractor
Yee and Sedivy (2006)	Target	**Semantic**	Distractor	Distractor
Huettig and Altmann (2005)	**Semantic**	Distractor	Distractor	Distractor
Mirman and Magnuson (2009)[a]	Target	**Near Sem**	**Far Sem**	Distractor
Huettig and McQueen (2007)[b]	**Onset**	**Semantic**	**Visual**	Distractor

The items displayed within scenes in each empirical study are listed with observed competitor effects highlighted in bold. Visual = visual competitor, Semantic = semantic competitor, Onset = phonological onset competitor, Rhyme = phonological rhyme competitor.
[a] Study presented near and far semantic competitors on separate trials.
[b] Experiment 1.

semantic effects observed in the visual world paradigm (Huettig, Quinlan, McDonald, & Altmann, 2006; Mirman & Magnuson, 2009; cf. Yee & Sedivy, 2006) in that it fixates items that share a semantic relationship at levels proportional to the number of semantic features shared between items. In contrast fixation of items that share purely phonological relationships is, in addition, dependent on the temporal location of overlapping phonological features. The model also replicates the difference between effects of phonological rhyme and phonological onset overlap as reported in Allopenna, Magnuson, and Tanenhaus (1998), fixating items that share initial phonemes earlier and with increased probability than items that share phonemes in final positions.

This model of language mediated visual attention has been shown to be not only sufficient to account for the experimental effects of the visual world paradigm for literate participants but also, importantly for this investigation, demonstrates how such features of behaviour are emergent properties of both the structure of representations and the computational properties of the mappings performed between them. The model achieves this through capturing both the process through which this behaviour is acquired and through use of a parsimonious architecture that implements only minimal assumptions about processing mechanisms. The model therefore is appropriate for testing the impact of differences between populations in representational structure on eye gaze in the visual world paradigm.

Previous computational models of the effects of phonological processing used in conjunction with visual world data have tended to model processing in a single modality (Allopenna et al., 1998; Mirman, Dixon, & Magnuson, 2008), the model used in this study however captures the processing of phonological, semantic and visual information. The model provides as an output a dynamic measure of the location of fixation across multi-object scenes over time, which is dependent on the integrated processing of information across all three modalities. The model also differs from previous models of effects on phonological processing where phonology has been used as both an input and output measure (Harm & Seidenberg, 1999). In contrast the chosen model has a different dependent variable 'eye gaze' therefore phonological manipulations are only indirectly related to behaviour. The ability to detect simultaneously the effect of phonological and semantic influences on performance permit greater discrimination of the effects of the phonological representation or the cognitive efficiency theory of literacy by investigating whether each implemented theory matches the effects of literacy in both representational domains.

We manipulate both grain-size in phonological processing as well as processing efficiency within this neural network model of language mediated eye gaze (Smith et al., 2013a,b). As finer grained representations more clearly encode the regions of representations that differ or overlap, we predict that increasing the granularity of phonological processing will increase the salience of phonological onset competitors at points in which the phonology of the unfolding spoken word overlaps with the phonological representation corresponding to the phonological onset competitor and reduce its salience at points in which the signal mismatches. Therefore a system that processes phonological information at a finer grain size will be able to use such information to distribute attention more dynamically in response to information in the unfolding auditory input.

We also predict that semantic effects on the other hand should not be affected significantly by the granularity of phonological processing. In previous computational simulations of language mediated eye gaze (Smith et al., 2013a,b), semantic effects are driven by overlap between the semantic representations of the visually depicted object and the spoken word. Overlap effects are therefore dependent on the level of activation of overlapping semantic features triggered by the concurrent visual and auditory input. As more of the phonological signal unfolds, the activation of corresponding semantic features will increase. When the words' semantic properties are maximally activated, if the level of representational overlap does not differ across distinct phonological grain sizes, then the literate and illiterate simulations should activate semantic representations equally.

In addition, we predict that manipulations of processing efficiency will have a greater impact on semantic effects than phonological effects. Within the model (see Section 2.1), activation of semantic information is more sensitive to the efficiency of information transfer within the network as it is not directly activated by the visual or auditory input, but instead activated as a consequence of signals that flow through the network from phonological and visual input layers. We also predict that such effects will be quantitative rather than qualitative in nature as the structure of signal overlap will not differ but simply lead to an overall reduction in the activation of overlapping features, which in turn will result in a quantitative reduction in the saliency of semantic competitors.

2. Method

2.1. Architecture

The neural network model used within this paper is based on the ASR model of language mediated eye gaze presented in Smith et al. (2013a). The same network architecture (see Fig. 2) was used for all simulations. The model consisted of four modality specific processing layers connected via a central resource. We know from behavioural data recorded in visual world studies that language mediated visual attention is driven by the interaction of information extracted from the visual environment and speech signal in terms of semantic, visual and phonological representations (e.g., Huettig &

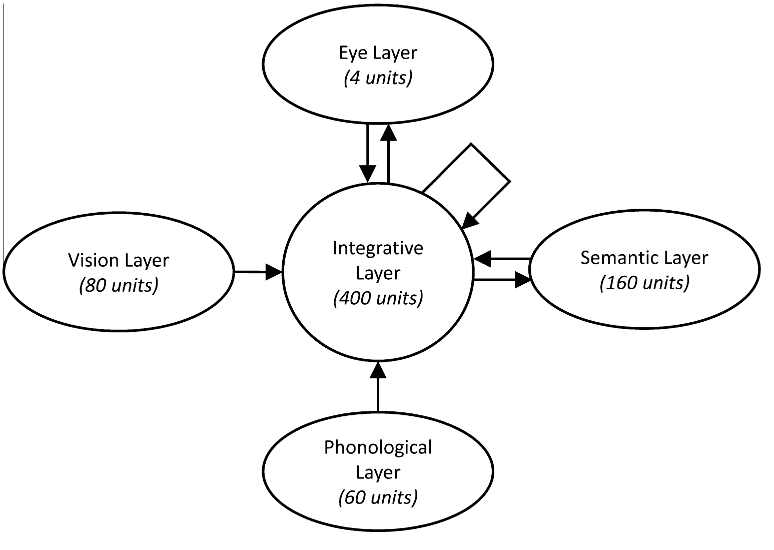

Fig. 2. The ASR Model of language mediated visual attention.

McQueen, 2007). The ASR architecture offers a parsimonious solution to how these modalities interact, and a means by which the emergent properties of this complex interaction can be harnessed. The architecture allows competition at multiple levels of representation, parallel activation of representations, the integration of information from multiple modalities and allows for both inhibitory and excitatory associations between representations, all of which have been proposed as important theoretically for reflecting behavioural effects in the Visual World Paradigm (see Smith et al., 2013a, for review). The philosophy of the model was to investigate the extent to which information from different modalities interacts in language processing tasks, with constraints within the model's processing resulting from the nature of the mappings between representations, rather than imposed architectural assumptions.

The visual layer (80 units) simulates the extraction of visual information from up to four locations in the visual field. The layer is divided into four 20 unit slots. Each slot encodes the information available at a single location in the visual field, with active units representing visual feature information about objects. The phonological layer (60 units) simulates the temporally unfolding speech signal. It comprises six time slots, each of which contains 10 units which encode the phonological features of the auditory input at a given time point. Phonological layer time slots are activated sequentially during the presentation of the spoken word to the model. For example, at word onset the activation pattern corresponding to the initial information of the unfolding spoken word will be presented to the first 10 units in the phonological layer, while all other units in the layer remain inactive. At the next time step the activation pattern corresponding to the second portion of the spoken word will in addition be presented to units 11–20 in the phonological layer, while again all subsequent units will remain inactive. This process continues, with an additional phonological portion presented at each subsequent time step until the entire phonological representation of the spoken word is presented to the phonological layer (i.e. all six portions corresponding to the six time slots). The semantic layer (160 units) represents the semantic features for items presented either in spoken or visual object form. Finally, the eye layer (4 units) provides a measure of the model's direction of gaze across the four possible locations in the visual field, with each unit in the eye layer associated with one of the four quadrant locations in the visual field. Activation of eye layer units was taken to correspond to the probability of fixating associated locations in the visual environment, with the location associated with the most highly activated eye layer unit interpreted as the location currently fixated. All four layers were fully connected to a central integrative layer (400 units) which was fully self-connected and also fully connected to allow activation to feed back to eye and semantic layers.

At each time step activation passed between all layers in the network. Training trials extend over a total of 14 time steps to enable activation to cycle between representations in the model. During testing, this period was extended to allow for insight into the time-course of interaction between representations across modalities within the model.

2.2. Representations

To ensure the overlap for visual, phonological, and semantic representations was fully controlled both within and across modalities, a fundamentalist approach was taken in their construction, i.e., that "a model should embody only the principles that are theorized to account for the phenomenon in focus" (Kello & Plaut, 2000). Consequently, artificial corpora were constructed each consisting of 200 items (or words in the artificial corpus), with each item defined by a unique visual, a unique semantic and a unique phonological representation. We constructed 15 versions of the corpus for each manipulation of grain size, though corpora were matched across cognitive efficiency manipulations, resulting in 45 artificial corpora (3 grain sizes $\times$ 15 instantiations of each grain size). This was to ensure that chance variation in distribution of the spoken forms of words, and chance random starting states of the model, did not bias model performance.

Visual representations were 20 unit binary vectors with each unit representing the presence or absence of a distinct visual feature. Visual representations were composed of two distinct components, one component representing coarse (low frequency) visual features and the other fine (high frequency) visual features. Visual features were randomly assigned to items with $p(\text{active}) = 0.5$. Ten units per component meant that encoding of 200 unique representations was possible without populating the representational space to maximal density, ensuring that visual representations could be distinguished by one or more visual features.

Semantic representations were designed to reflect discrete and relatively sparse semantic feature-based representations for words (Harm & Seidenberg, 2004). For each item a unique set of 8 from a possible 200 semantic features were activated, with features randomly assigned.

Representational overlap between items was controlled in visual, semantic and phonological dimensions (see Table 2). Embedded within each corpus were 40 target items. Within the corpus were also embedded for each target item a semantic competitor ($n = 40$) and a phonological competitor ($n = 40$). Competitors shared increased representational overlap with their assigned target item in a single modality (either semantic or phonological). Semantic competitors shared 4 of 8 semantic properties with their assigned target, while all unrelated items shared a maximum of 1 semantic property with any other item. For a definition of phonological competitors see Section 2.3.

2.3. Simulating differences in phonological processing

To simulate differences in the granularity of speech processing three forms of processing of the phonological input were constructed: fine, medium, and coarse grained phonological representations, reflecting variations in the componentiality of the phonology, from single phonemes up to a holistic word-level representation. Fine grained representations simulated phonological processing in which the unfolding speech signal activates a componential sequence of phonemes, and this was the representation used in previous simulations of standard speech processing with the ASR model (Smith et al., 2013a,b). For the fine grained simulations, an inventory of 20 possible phonemes was constructed. Each phoneme was encoded as a unique 10 unit binary vector with units assigned with $p(\text{active}) = 0.5$, and each unit representing a phonological feature. For each word, a unique sequence of 6 phonemes was constructed by randomly sampling from the set of 20 phonemes. For the overlapping target-competitor pairs, the target and competitor shared their initial 2 phonemes (in Huettig et al., 2011 phonological competitors shared at least the first two phonemes with the target word). No other word pairs within the corpus shared their initial two phonemes and no two words shared their initial three phonemes. These constraints applied equally to all the grain size conditions in order to ensure that comparisons between the grain sizes were controlled. However, the overall amount of overlap between items in the whole corpus is less than that observed in natural language, where overlap even up to the first 5 phonemes of 6 phoneme words is observed. This minimal overlap may have enhanced the size, rather than the qualitative nature, of phonological overlap observed in the results. All items had a unique sequence of the final three phonemes and no item contained more than two of the same phonemes in its entire representation.

Table 2
Cosine distance between target and distractor representations. Records mean ($\bar{x}$) and standard deviation [σ] calculated across all items and corpora (15 per grain size).

Modality	Simulation	Distractor	Cosine distance ($\bar{x}$, (σ))		
			Onset	Rhyme	Overall
Phonological					
	Fine grain	Competitor	0.00 (0.00)	0.51 (0.12)	0.34 (0.08)
		Unrelated	0.51 (0.16)	0.49 (0.11)	0.50 (0.09)
	Moderate grain	Competitor	0.17 (0.10)	0.42 (0.09)	0.33 (0.07)
		Unrelated	0.51 (0.14)	0.51 (0.10)	0.51 (0.08)
	Coarse grain	Competitor	0.35 (0.12)	0.34 (0.09)	0.34 (0.07)
		Unrelated	0.51 (0.14)	0.50 (0.10)	0.50 (0.08)
			Overall		
Semantic					
	All simulations	Competitor	0.50 (0.00)		
		Unrelated	0.96 (0.06)		

For moderate grained representations two components encoded each word, analogous to syllable level representation of the stimulus. Each component was a unique 30 unit binary vector with units assigned with $p(\text{active}) = 0.5$. For the phonological overlap items, the target and competitor shared 2/3 of the features present in the initial component, and the remaining features in both the initial and second component were either shared or distinct with $p = 0.5$. The number of overlapping features was thus controlled across the fine and medium grained representations for phonologically overlapping items (see Table 2), though they differed in terms of the componentiality of the overlap. Unrelated items shared all features with $p = 0.5$. Phonological similarity was thus simulated as distributed over the first half of the item, rather than, as in the case of the fine grained simulations, with precise similarity over the first two phonemes of the item.

For coarse grained representations each word was defined by a single component, simulating a word-level phonological representation. This component was a unique 60 unit binary feature vector with $p(\text{active}) = 0.5$. For the overlapping pairs of target and competitor items, 1/3 of the features were identical between the two items, and the remaining features were shared with $p = 0.5$. Thus, the total number of similar features was the same as for the overlapping items in the fine and medium grained representations. Unrelated items shared features with $p = 0.5$.

To simulate the time-course of the unfolding phonological input, for all simulations sequences of 10 features of the spoken input were gradually presented. Thus from word onset an additional 10 features of the spoken input were presented at each subsequent time step, until by word onset +5 time steps, all 60 features that define an item's spoken representation were presented as input to the phonological layer. This controlled for the temporal presentation of the auditory signal for each grain-size encoding, but the grain size of representations differed in terms of the componentiality of the presented features. Thus, for the fine grained representations, the presentation was phoneme-by-phoneme, whereas for the moderate grained representation, the presentation partially unfolded the syllable across three time steps, such that after three time steps the model was exposed to a full component. However, early stages of presentation did provide the model with information about the general sound of the syllable (so two similar syllables would have similar representations in the first ten features). For the coarse grained representation, again the presentation partially unfolded the word across six time steps, with similar sounding words having similar representations as the spoken form unfolded. Although, during language processing listeners would receive the same auditory signal with identical temporal properties, the grain size at which an individual processes speech will determine their ability to identify the components of words. Hence, when an onset competitor shares its initial two phonemes with a spoken target word (but not the entire syllable) a system processing at a finer grain will be quicker to detect the speech sound overlap than a system processing at a coarser grain. The fine, moderate and coarse grain representations implemented within this model capture this assumption.

2.4. Simulating differences in cognitive efficiency

There are several ways to simulate variation in cognitive efficiency in neural network models, including addition of noise to activations of units, reductions in processing resources to form mappings between representations, or reduction in overall levels of activation within the network (Harm & Seidenberg, 1999; Monaghan & Shillcock, 2004). We chose to simulate cognitive efficiency in the current models in terms of addition of noise. This decision was made so as to link the implementation of cognitive efficiency as closely as possible to theories about the information processing effects of literacy on neural processing, in terms of myelination of highly-trained pathways reducing noise levels in neural transmission via faster processing of high quality information (see Section 1.2). Gaussian noise was thus applied $[N(\mu = 0, \sigma^2 = 0.02)]$ to the output of all units in the network for the lower cognitive efficiency simulation, but no noise was added to the higher cognitive efficiency version.[1] This

[1] The cognitive efficiency hypothesis as implemented within this study does not argue for a reduction of cognitive efficiency within the entire cognitive system. Instead, based on the argument of increased myelination in heavily trained networks, differences in cognitive efficiency would only result in networks that experience a difference in levels of training as a consequence of literacy acquisition. Within the model we only model such networks and therefore manipulation of global cognitive efficiency within the model is a valid reflection of the neural network changes associated with literacy.

resulted in differences in the fidelity with which information passed through the network, and consequently the speed at which activation could accumulate in different modality layers in the model. Pilot simulations were used to establish an appropriate level of noise for the lower cognitive efficiency simulation. Simulations trained with noise sampled from [$N(\mu = 0, \sigma^2 = 0.05)$] failed to learn some of the mappings between modalities that were a precursor to testing experimental performance of the model against the behavioural data (see Section 2.5) and simulations trained with noise sampled from [$N(\mu = 0, \sigma^2 = 0.01)$] displayed negligible differences in performance from simulations in which no noise was applied.

For each simulation set (fine grain, low efficiency; fine grain, high efficiency; moderate grain, low efficiency; moderate grain, high efficiency; coarse grain, low efficiency; coarse grain, high efficiency), 15 versions were trained on one of the distinct corpora and each was initialised with a different random seed.

2.5. Training

For each grain size (fine, medium, and coarse) we manipulated the model's cognitive efficiency (high, low) leading to a total of six sets of parameters for the model, and as mentioned in Section 2.2, there were 15 different simulation runs for each of these parameterisations.

All simulations were trained on four tasks (see Table 3) that aimed to simulate the tasks performed by participants in the natural learning environment. We assume that participants gain knowledge of an item's visual, semantic and spoken form by repeated and simultaneous exposure to these multiple forms of representation: It is through such experience that individuals acquire the associations between representations across modalities that later drive the behaviour observed in the laboratory setting.

2.5.1. Vision to semantics

This task aimed to simulate the learning that occurs during events in which individuals simultaneously view an item and determine some of its semantic properties, e.g., its function: seeing a fork and determining its use for eating. This was simulated by first randomly selecting four items from the artificial corpus, one of which was randomly selected as a target. The visual representations of each of the four items was then presented to the model at trial onset (time step 0), with each item randomly assigned to one of the four locations in the model's visual field. The eye unit relating to the location of the target's visual representation was also fully activated at trial onset with all other eye units fixed at zero activation. These values remained fixed for the remainder of the training trial. Throughout the trial small levels of variable background noise were provided as input to the phonological layer, simulating ambient background sound. Once sufficient time had passed allowing for activation to flow from the visual and eye layers to the semantic layer (i.e. time step 3) the item's semantic representation was provided as a target and error was backpropagated through the network up to time step 14.

2.5.2. Phonology to semantics

This aimed to simulate the learning that takes place when an individual is exposed to an item's spoken form and is required to determine its semantic properties, for example, when hearing the word fork and eating from a fork. To simulate such occurrences, an item was first randomly selected from the training corpus and assigned the role of target. At trial onset the first 10 features of its phonological representation were presented in the initial slot of the phonological layer. At each subsequent time point a further 10 features of the target's representation were presented in the corresponding phonological input layer slots until the entire representation had unfolded. This remained present until the end of the training trial. Throughout such trials, random background noise was presented to the visual layer to simulate ambient stimulation of the visual system. Once the entire word had unfolded and sufficient time had elapsed for a signal discriminating the target from possible competitors to pass to the semantic layer (time step 5), the item's semantic representation was presented as a target and error was backpropagated until time step 14.

Table 3
Procedure for the model's training trials.

Task	Vision		Phonological		Semantic		Eye	
	Description	Time step	Description	Time step	Description	Time step	Description	Time step
1. Vision to Semantics	4 visual representations randomly selected from the training corpus, 1 of which is randomly selected as a target	0–14	Random time variant noise provided as input	0–14	Target's semantic representation provided	3–14	Target location fully activated, all other locations inactive	0–14
2. Phonological to Semantics	Random time invariant noise presented to all visual input slots	0–14	Target speech signal provided as staggered input	0–14	Target's semantic representation provided	5–14	No constraints on activation	
3. Phonological to Location	Identical to procedure in task 1 and 4	0–14	Identical to procedure in task 2	0–14	No Constraints on activation		Target location fully activated, all other locations inactive	5–14
4. Semantics to Location	Identical to procedure in task 1 and 3	0–14	Identical procedure in task 1	0–14	Semantic representation of target provided	0–14	Target location fully activated, all other locations inactive	2–14

2.5.3. Phonology to vision

Orientating to an item when hearing its spoken form was trained by first selecting four items randomly from the training corpus and selecting one as a target. The visual representation of all four items was presented to the visual input layer at trial onset with locations in the visual field randomly assigned. Also coinciding with trial onset, the phonological representation of the target item began to unfold, with an additional 10 features of the target's phonological representation presented at each subsequent time step. Once the entire word had unfolded and sufficient time had passed to allow a discriminating signal to reach the eye layer (time step 5), the training signal was provided. This consisted of fully activating the eye unit relating to the location of the target's visual representation while fixing activity in all other eye units to zero.

Semantics to vision. A similar procedure was applied when training the model to orientate to the location of a target when provided with its semantic representation. Again four items were randomly selected from the training corpus and one randomly assigned as the target. The visual representations of all four items were presented to the visual input layer at trial onset, with locations randomly assigned. Also coinciding with trial onset, the semantic representation of the target was presented. Throughout such training trials small levels of variable noise were provided as input to the phonological layer to simulate auditory background noise. Once sufficient time had elapsed for the signal from both visual and semantic layers to pass to eye layer units (time step 2), the training target signal was provided. This consisted of fully activating the eye layer unit associated with the location of the target's visual representation with zero activation in all other eye layer units.

All training tasks were randomly interleaved. In the natural language learning environment, items around the child are frequently left unnamed (Yu & Ballard, 2007). Hence, we assume during training that items based on their semantic properties are selected more frequently than items based on their spoken form, and so phonology to vision tasks were four times less likely to occur in training than other training tasks.

Connection weights were initialised with random weights taken from a uniform distribution [−0.1, 0.1]. Weights were adjusted online during the training process using recurrent back-propagation with learning rate 0.05. All simulations were trained on 850,000 training trials as this provided sufficient exposure for simulations to perform accurately on all four training tasks.

3. Results

3.1. Pre-test

Post-training, all simulations were tested on their ability to perform each of the four training tasks. Table 4 presents the accuracy of simulations on each task, averaged across 15 instantiations of each simulation. For tasks requiring the model to reproduce the semantic representation of the target when presented with its visual representation, simulations were tested on their ability to perform this task with the target presented in all possible locations in the visual field. Similarly, on orientation tasks, simulations were tested on their ability to orientate to the location of the target when the target was positioned in each of the four possible locations in the visual field. For phonological to semantic mapping tasks the model was tested four times on each item. Two measures were used to assess performance on each training task. For semantic mapping tasks we calculated the cosine distance between mean activation in the semantic layer during test trials and the semantic representation of all items within the training corpus. In Table 4, mean accuracy indicates the proportion of test trials for which the target's semantic representation was closest in terms of cosine distance to activation within the semantic layer. The second measure indicates the proportion of items within the training corpus for which activation in the semantic layer was closest to the semantic representation of the target in at least three out of four test trials. Two measures were also collected to assess model performance on orientation tasks. Mean accuracy indicates the proportion of test trials in which the eye unit relating to the location of the target's visual representation was most highly activated. The second measure provides the proportion of items within the training corpus for which the eye unit relating to the location of the target's visual representation was most highly activated on 3 out of 4 test trials.

The measures of model performance presented in Table 4 demonstrate that all simulations were able to complete each of the four training tasks with a high degree of accuracy and displayed comparable levels of performance. There are significant differences between high and low cognitive efficiency simulations in task performance for visual to semantic mappings for the mean activation but not the % items measure, indicating that cognitive efficiency drives a small difference in mapping location to semantic representations, as was predicted, but does not affect mapping from semantics, nor mappings from other modalities to semantics.

3.2. Simulating the effects of grain size and cognitive efficiency on language mediated visual attention

To simulate the conditions under which participants were tested in Huettig et al. (2011), Experiment 1, we presented the model with scenes containing a semantic competitor, a phonological onset

Table 4
Trained model's performance on training tasks.

Simulation		Task							
Efficiency	Grain	Visual to semantic[b]		Phonological to semantic[b]		Phonological to location[a]		Semantic to location[a]	
		Mean	% items	Mean	% items	Mean	% items	Mean	% items
High	Fine	0.88	0.98	1.00	1.00	0.91	0.96	0.92	0.98
Low	Fine	0.85	0.97	1.00	1.00	0.90	0.97	0.92	0.99
High	Moderate	0.86	0.97	1.00	1.00	0.91	0.97	0.92	0.98
Low	Moderate	0.83	0.97	1.00	1.00	0.90	0.96	0.93	0.99
High	Coarse	0.87	0.98	1.00	1.00	0.91	0.97	0.93	0.99
Low	Coarse	0.82	0.97	1.00	1.00	0.90	0.96	0.93	0.99

[a] *Location task:* mean = mean proportion of test trials in which eye unit corresponding to target location was most highly activated, % items = proportion of items within corpus for which target location was most highly activated on at least 3 of 4 test trials;

[b] *Semantic task:* mean = proportion of test trials for which semantic layer activation was closest to target's semantic representation, % items = proportion of items for which semantic layer activation is closest to target's semantic representation on at least 3 of 4 test trials.

competitor and two unrelated distractors while simultaneously presenting the phonological representation of the given target word. This was achieved using the following procedure. At trial onset the visual representations of the four items within the scene were presented. After a short delay, to enable pre-processing of the visual information, the spoken target word began to unfold (time step = 5), with an additional component of the target word's phonological representation presented at each subsequent time step until the full representation was revealed. The model's "gaze", was interpreted as being directed towards the location in the visual display associated with the most highly activated unit in the eye layer. 'Gaze' was recorded in this manner at each time step throughout the test trial, which lasted in total 29 time steps. For each instantiation of each simulation there were in total 960 test trials, with each item ($n = 40$) occurring with competitors in all possible spatial configurations ($n = 24$). Fig. 3 displays the change in $p(\text{fix})$ from word onset (time step = 5), for each simulation (Fig. 3A: High cognitive efficiency simulations; Fig. 3B: Low cognitive efficiency simulations), with $p(\text{fix})$ recording the Luce ratio of fixations for each category of item within displays, averaged over all test trials ($n = 960$) and instantiations ($n = 15$).

We used linear mixed effects models (Baayen, Davidson, & Bates, 2008; Barr, 2008; Jaeger, 2008) to analyse the influence of grain size and cognitive efficiency on differences between competitor and distractor fixation. As a baseline for behaviour, we identified a preview time window (time steps 0–7), the period from trial onset until the first time point in which a signal discriminating between target and competitor within the phonological input can influence eye layer units. We then compared this baseline eye gaze performance of the model to its behaviour in time windows after this point, where the information from the phonological input met the visual system, distinguishing between early (time steps 8–18), and late (time steps 19–29) processing.

For each time window in each test trial we calculated the total number of time steps a given category of item was fixated, *fix*(*item category*). As two unrelated distractors were present in each scene, for each time window (preview [time step 0–7], early [8–18], late [19–29], or early & late [8–29]) we divided the total number of fixations towards unrelated distractors by two. These totals were then used to calculate the empirical logits (log odds were used to avoid issues arising from calculating estimates based on proportion data, see Jaeger, 2008) for each category of item within each time window. We then calculated the difference between the log-odds of fixating a given competitor type and the log-odds of fixating unrelated distractors. This difference (*competitor bias*, see Eq. (1)) formed our dependent measure as it reflects the difference in fixation behaviour as a consequence of representational overlap. Separate analyses were conducted for each competitor type (semantic competitor – unrelated distractor [semantic bias], and phonological competitor – unrelated distractor [phonological bias]).

Our initial analysis compared across simulations to examine whether the difference between fixation of competitor and distractor altered between the preview period and the period post target word onset (interest window = time steps 8–29), and whether this difference was influenced by either decreases in the granularity of phonological processing (i.e. whether behaviour in coarse and moderate grain size simulations differed from fine grain size simulations) or as a consequence of differences in cognitive efficiency. To compare behaviour between simulations we predicted the dependent measures with fixed effects: phonological grain size (coarse, moderate or fine: with fine grain mapped onto the intercept forming the baseline condition), cognitive efficiency (coded as a numerical factor centred on zero: high = −0.5, low = 0.5) and time window (coded as a numerical factor centred on zero: preview = −0.5, interest window = 0.5). The random effects structure included random intercepts for both instantiation and item as well as random slopes for time window both by instantiation and by item. This is the maximal random effect structure (Barr, Levy, Scheepers, & Tily, 2013), as both grain and noise were varied between instantiation and item. To derive p-values we assumed t-values were drawn from a normal distribution (Barr, 2008).

$$\text{competitor bias} = \log\left(\frac{\text{fix(competitor)} + 0.5}{\text{fix(competitor)} - \text{total time steps} + 0.5}\right) - \log\left(\frac{\text{fix(distractor)} + 0.5}{\text{fix(distractor)} - \text{total time steps} + 0.5}\right) \tag{1}$$

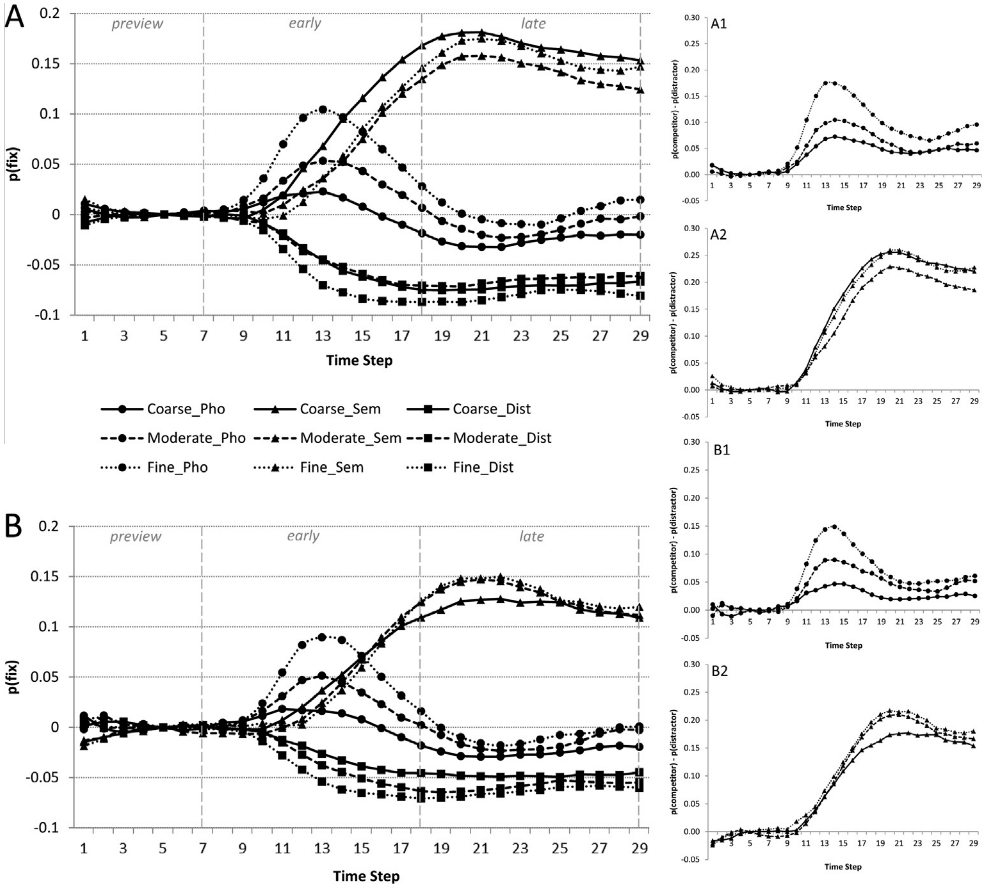

Fig. 3. Time course of fixation behaviour displayed by the ASR Model. Figures A and B display the proportion of fixations [p(fix)] directed towards items within scenes containing a phonological onset competitor (Pho), a semantic competitor (Sem) and two unrelated distractors (Dist). Fixation proportions are plotted for fine grain (Fine), moderate grain (Moderate) and coarse grain (Coarse) simulations with high cognitive efficiency (Fig. 3A) and low cognitive efficiency (Fig. 3B) (compare to behavioural data in Fig. 1A and B). Figure A1 and B1 display the difference in fixation of phonological competitors compared to distractors in the high cognitive efficiency condition and low cognitive efficiency condition respectively, while figures A2 and B2 display the difference in fixation of semantic competitors compared to distractors in the high cognitive efficiency condition and low cognitive efficiency condition respectively.

Results of this analysis showed that simulations displayed an increased *phonological bias* post word onset ($\beta = 0.454$, SE = 0.056, $t = 8.064$, $p < 0.001$). Coarse grain simulations differed from fine grain simulations displaying a reduced *phonological bias* ($\beta = -0.202$, SE = 0.036, $t = -5.551$, $p < 0.001$), importantly there was also an interaction between grain size (fine compared to coarse) and time window showing that fine grain simulations compared to coarse grain simulations displayed a further increased *phonological bias* post word onset ($\beta = -0.338$, SE = 0.073, $t = -4.646$, $p < 0.001$). Fine grain simulations also displayed an increased *phonological bias* compared to moderate grain simulations but only in the period post word onset ($\beta = -0.182$, SE = 0.073, $t = -2.509$, $p = 0.012$). There was no effect of cognitive efficiency on *phonological bias*.

Applying the same analysis to examine *semantic bias* showed that simulations displayed an increased *semantic bias* post word onset ($\beta = 0.853$, SE = 0.060, $t = 14.242$, $p < 0.001$). There was no difference in *semantic bias* between moderate and fine grain simulations. The only significant difference between simulations of differing grain size was in the significant interaction for coarse compared to fine grain and cognitive efficiency ($\beta = -0.167$, SE = 0.074, $t = -2.248$, $p = 0.025$). This shows that reductions in cognitive efficiency lead to a lower *semantic bias* in coarse grain simulations compared to fine grain simulations. There was also an overall marginally significant increased *semantic bias* as a result of reductions in cognitive efficiency ($\beta = 0.100$, SE = 0.052, $t = 1.905$, $p = 0.057$).

We also applied the same approach to examine differences between simulations over the three theoretically motivated time regions (preview, early, and late), as previous research suggests that processing may differ in how effects are distributed across time, post word onset (e.g. Huettig & McQueen, 2007). This was done by splitting the remainder of the test trial, post preview, equally into two further windows, an early window (time step 8–18) and a late window (time step 19–29). The model structure used for analysis was identical to previous analyses except that we also included the additional window in the fixed effect of time window (preview, early or late: with preview mapped onto the intercept forming the baseline condition).

Results showed that simulations displayed a significant increase in *phonological bias* in the early versus preview window ($\beta = 0.531$, SE = 0.075, $t = 7.048$, $p < 0.001$). This effect was only marginally significant in the late window ($\beta = 0.148$, SE = 0.085, $t = 1.737$, $p = 0.082$). A positive yet lower parameter estimate in late over early windows suggests a reduced bias towards fixating phonological competitors in later windows. We did not observe a main effect of either coarse or moderate grain. However, there was an interaction between coarse compared to fine grain and early versus preview window ($\beta = -0.411$, SE = 0.098, $t = -4.202$, $p < 0.001$), and coarse compared to fine grain and late versus preview window ($\beta = -0.300$, SE = 0.098, $t = -3.063$, $p = 0.002$). This means that although coarse and fine grain simulations did not differ in the preview period they did differ in both the early and late window. Parameter estimates indicate that the phonological effect was stronger in the fine grain simulation than the coarse grain simulation in both early and late windows, although a lower parameter estimate suggests this difference was lower in the late window. An interaction between moderate versus fine grain and early versus preview window ($\beta = -0.230$, SE = 0.098, $t = -2.351$, $p = 0.019$) was also observed, however there was no interaction between moderate versus fine grain and late versus preview window. This shows that the phonological effect only differed between moderate and fine grain simulations in the early window. A negative parameter estimate suggests fine grain simulations displayed a greater phonological effect in this window than moderate grain simulations. Thus, cognitive efficiency had no influence on *phonological bias*, with no significant main effect nor interactions.

Applying the same model structure to predict *semantic bias* yielded a main effect for both early ($\beta = 0.499$, SE = 0.073, $t = 6.877$, $p < 0.001$) and late windows ($\beta = 1.261$, SE = 0.087, $t = 14.475$, $p = 0.001$). Increasing positive parameter estimates suggest an increasing bias toward fixating semantic competitors in early and late windows over preview periods, with a greater bias displayed in late over early windows. No other model parameters were significant.

We also examined the effects displayed by coarse, moderate and fine grain simulations individually. For each grain size we again used mixed effects models to predict measures of *semantic bias* and *phonological bias*. Models included fixed effects of time window (preview, early or late: with preview mapped onto the intercept forming the baseline condition) and cognitive efficiency (coded as a numerical factor centred on zero: high = -0.5, low = 0.5), in addition to a random effects structure that included random intercepts for both instantiation and item, and random slopes for time window both by instantiation and item (maximal random effects structure, Barr et al., 2013).

Fine grain simulations displayed increased fixation of phonological competitors over unrelated distractors in the early window compared to the preview window ($\beta = 0.531$, SE = 0.075, $t = 7.081$, $p < 0.001$), yet no difference was observed in this measure in the late window when compared to preview. Moderate grain simulations displayed a similar pattern of behaviour with an increased bias towards fixating phonological competitors in early windows over preview ($\beta = 0.301$, SE = 0.071, $t = 4.228$, $p < 0.001$), yet no difference between late windows and preview. Coarse grain simulations however, did not display a significant difference in *phonological bias* in either early or late windows when compared to preview. There was no evidence for an influence of cognitive efficiency on *phonological bias* for coarse, fine or moderate simulations.

Fine, coarse and moderate grain size simulations all displayed an increased bias towards fixating semantic competitors over distractors in both early and late windows when compared to the baseline preview window (fine [early: $\beta = 0.499$, SE = 0.076, $t = 6.596$, $p < 0.001$; late: $\beta = 1.261$, SE = 0.106,

$t = 11.880$, $p < 0.001$], moderate [early: $\beta = 0.450$, SE = 0.073, $t = 6.178$, $p < 0.001$; late: $\beta = 1.196$, SE = 0.103, $t = 11.666$, $p < 0.001$]; coarse [early: $\beta = 0.474$, SE = 0.085, $t = 5.586$, $p < 0.001$; late: $\beta = 1.184$, SE = 0.120, $t = 9.865$, $p < 0.001$]). There was no evidence for an influence of cognitive efficiency on *semantic bias* in either the moderate or fine grain simulations. However, in the case of coarse grain simulations there was a significant interaction between cognitive efficiency and the late versus early time window ($\beta = -0.386$, SE = 0.138, $t = -2.786$, $p = 0.005$). A negative parameter estimate indicates that low cognitive efficiency leads to a reduction in the magnitude of the semantic effect in the late window over preview, for coarse grain simulations.

A further post hoc analysis was also conducted to examine whether there was any evidence in the fixation behaviour of coarse grain simulations, post word onset, for an effect of phonological overlap. The phonological competitor bias in the preview period was compared to the same measure aggregated over all time steps in both early and late windows (cf. Huettig et al., 2011, Experiment 2). The model used for this analysis was identical to that used previously to examine effects at a single grain size, the only difference being to the fixed effect of time window (coded as a numerical factor centred on zero: preview = -0.5, combined early and late window = 0.5). This analysis showed that coarse grain simulations displayed a marginally significant increased bias in fixating phonological competitors over distractors post word onset ($\beta = 0.116$, SE = 0.060, $t = 1.942$, $p = 0.052$). There was no evidence from this analysis for an effect of cognitive efficiency ($\beta = -0.052$, $t = -0.497$, $p = 0.619$).

In summary, both moderate and fine grain simulations displayed a phonological effect limited to early periods post word onset, with fine grain simulations displaying an increased *phonological bias* over moderate grain simulations in this period. Coarse grain simulations in contrast displayed a marginal increase in fixation of phonological competitors over unrelated distractors only when aggregating fixation across all time windows post word onset. Compared to fine grain simulations coarse grain simulations displayed a reduced *phonological bias* in both early and late windows post word onset. There was no evidence for an effect of cognitive efficiency on *phonological bias*.

Conversely, a marginal effect of cognitive efficiency was observed on *semantic bias* but only when pooling fixation behaviour across all time windows post word onset. All simulations displayed an increased *semantic bias* in early and late windows. The only difference between simulations of differing granularity in their *semantic bias* was observed in the interaction between coarse versus fine grain and cognitive efficiency. Also coarse grain simulations displayed a reduced *semantic bias* in late windows as a consequence of a reduction in cognitive efficiency.

4. Discussion

Our modelling results successfully replicated qualitative differences observed between high and low literates in sensitivity to phonological competitors, reported in Huettig et al. (2011). As was displayed in the behaviour of high literates, fixation of phonological onset competitors by fine and moderate grain size simulations was closely time locked to overlap between the competitor's phonological representation and the unfolding speech signal. Fine and moderate grain size simulations displayed an initial bias towards fixating phonological competitors shortly after word onset, and a rapid decline in fixation and return to baseline distractor levels once the phonological signal mismatched. Coarse grain simulations on the other hand displayed less dynamic changes in fixation of phonological competitors in response to overlap in the speech signal and fixated phonological competitors at levels close to unrelated distractors. Unlike fine and moderate simulations, for coarse grain simulations a marginal bias toward fixating phonological competitors over distractors was only observed when pooling fixation behaviour across both early and late windows. Such behaviour is similar to that displayed by low literates (Huettig et al., 2011): When presented with scenes containing semantic and phonological competitors low literates did not display a bias towards phonological competitors. When tested under more sensitive conditions, in which only phonological competitors were present, low literates did display sensitivity to phonological overlap and looked marginally more towards phonological competitors compared to distractors in the first 1000 ms post word onset. As in the case of the coarse grain size simulations, low literates did not display the rapid increase and decrease in looks towards phonological competitors in response to signal overlap as was shown by high literates.

Our simulations therefore demonstrate that differences in the granularity of phonological processing can modulate the phonological effect displayed in studies of language mediated visual attention, reflecting effects of literacy in the use made of phonological information in processing visual and semantic representations of stimuli. This feature of model behaviour was largely driven by the fact that more fine grained representations more precisely encode the regions of the word's phonological representation that differ or overlap. This information can then be exploited by the system to dynamically adjust fixation behaviour.

In contrast to the phonological effects, grain size did not modulate the magnitude of the semantic effect. In Huettig et al. (2011), differences in fixation of semantic competitors between high and low literates were only observed in later time windows, once the spoken word has unfolded. Within our simulations, fixation of semantic competitors is dependent on the level of activation of semantic properties shared by the semantic competitor and target. This activation is maximal when the entire phonological representation of the target has been input to the model. At this point, at the word level, representational overlap does not differ across grain sizes and therefore as the simulations show we do not observe a difference in semantic competitor fixations as a result of grain size manipulations.

For the cognitive efficiency manipulation in the model, it was predicted that phonological effects would be less evident than semantic effects following reductions in cognitive efficiency, because within the model, activation of semantic representations is entirely dependent on the efficiency of information transfer within the network, unlike phonological representations which are provided as a direct input. Pilot simulations demonstrated that increasing noise levels within the network (i.e. $N[\mu = 0, \sigma^2 = 0.05]$) were unlikely to lead to modulation of phonological effects, as this led to a failure in the model's ability to learn training tasks, tasks that we know both literates and illiterates are able to perform accurately. We also suggest that alternative methods of implementing cognitive efficiency within neural networks would also not simulate the pattern of the low-literate participants. For instance, reduction in the resources available for forming mappings between representations (Harm & Seidenberg, 1999) would impede the model's ability to learn tasks that both high and low literacy groups are capable of performing, such as vision to semantics, or phonological to semantics mappings. A further alternative implementation by reducing the overall levels of activation passing between layers within the model (Monaghan & Shillcock, 2004), or by increasing the threshold such that more activation is required before activating a response, is also unlikely to simulate the focused distinctions between high and low literate populations. Such an implementation would result in the same information being processed, just requiring longer in order to be processed. Thus, the same peaks of performance for the phonological competitors condition would be observed, but at a later point in time, and similar delays for all other mapping tasks that the model is required to perform. Therefore, phonological effects would still be observed yet would be delayed, along with delays to a broad range of other tasks, which is not the behaviour displayed by low literates. Although we accept that it is possible for other implementations of reductions in cognitive efficiency to have implications for processing beyond those captured by our simulations, our results are sufficient to indicate that reducing the quality of information transfer within networks was not adequate for explaining the qualitative difference in sensitivity to phonological overlap displayed by low literates in Huettig et al. (2011).

Also replicating the behaviour of both low and high literates, all simulations displayed an increasing bias toward fixating semantic competitors, across early and late windows, when compared to preview periods. Further, similar to the quantitative difference observed between high and low literates, the semantic bias was greater for fine grain simulations compared to coarse grain simulations with low cognitive efficiency. Analysis of the influence of cognitive efficiency on coarse grain simulations shows that, unlike fine and moderate simulations, who displayed no cognitive efficiency effects, low cognitive efficiency lead to a reduction in semantic bias in late windows. This is similar to the observed behaviour of low literates, as it is only in late windows that levels of semantic bias are significantly lower than those displayed by high literates, with a lower asymptote in fixation of semantic competitors for low literates. Semantic bias is dependent on the activation of semantic properties shared by both target and competitor. The strength of this effect is dependent on the level to which these units are activated by the visual input provided by the semantic competitor, and the phonological input from the spoken target word. Reducing the efficiency of information transfer within the network will reduce the strength of the signal travelling from visual and phonological layers, to activate associated semantic layer units, and hence the level of activation of overlapping semantic features. Unlike fine and moderate simulations, low cognitive efficiency only led to reduced semantic bias in coarse grain simulations. We suggest that this is most likely due to the componential structure of phonological representations in moderate and fine grain simulations that ensured activation of semantic features by phonological input was more robust to the introduction of noise. Importantly the differences between coarse and fine grain simulations captured the quantitative rather than qualitative differences in semantic bias observed between high and low literates. Within the model the effect of noise (cognitive efficiency) did not qualitatively alter the structure of information processed, but instead reduced the level to which overlapping semantic representations were activated and therefore semantic competitors fixated. Thus, it affected the role of representational overlap and the resulting fixation behaviour quantitatively rather than qualitatively.

Of the simulations conducted, only a comparison between a fine grain, high cognitive efficiency simulation and a coarse grain, low cognitive efficiency simulation replicated both the qualitative difference in sensitivity to phonological onset competitors and quantitative difference in sensitivity to semantic competitors observed between high and low literates in Huettig et al. (2011). Therefore, our simulations suggest differences in language mediated visual attention as a consequence of literacy training may well be driven by both mechanisms: changes to phonological encoding as well as increased cognitive efficiency. The model thereby presents the first explicit account of the consequences of literacy that extend beyond phonological processing to other aspects of language processing and highlights the necessity for multimodal computational models in order to gain insight into the inherently complex issue of multimodal interaction within human cognitive processing.

Although there is substantial evidence for an effect of literacy on speech processing, there have been very few computational modelling studies that focus on understanding the emergent consequences of training on orthographic mappings for phonological or semantic systems involved in speech processing. For instance, previous models of reading acquisition have made important contributions demonstrating an influence of orthographic transparency on phonological processing (Harm & Seidenberg, 1999; Yang, McCandliss, Shu, & Zevin, 2009) and semantic processing (Harm & Seidenberg, 2004; Yang, Shu, McCandliss, & Zevin, 2013). However, such models have tended to be trained on prototypical phonological representations in which substantial phonological structure is embedded, and then the processing of the phonological structure itself is investigated as both the input and output system. Such features of a model will have dramatic consequences for the type of structure to which the system develops sensitivity, and therefore these previous modelling studies are likely to have misrepresented the impact of orthographic training on the effects of the phonological grain size on processing within the language system more generally.

There remains a gap in our understanding of the extent to which literacy alters online speech processing (and broader aspects of cognitive processing) and the mechanisms through which it exerts an influence. To date the most influential cognitive models of speech processing do not allow for an influence of literacy on this process and have focused on modelling the behaviour of alphabetic literates (e.g. Distributed Cohort Model: Gaskell & Marslen-Wilson, 1997; TRACE: McClelland & Elman, 1986; Shortlist B: Norris & McQueen, 2008). We emphasise that a model of human speech processing should be sufficient to describe representations within the language system and their interaction, adequate for accounting for behaviour of literate and illiterate participants, as well as literates learning from different orthographies. The model presented in this paper does not simulate the emergent processes by which exposure to orthographic mappings leads to a restructuring of phonological representations or improved cognitive efficiency. It is possible that training on orthographic mappings also has emergent consequences for semantic processing that then give rise to increased sensitivity to semantic overlap. The empirical evidence suggests at most only a subtle effect of literacy on semantic processing (da Silva et al., 2004; Kosmidis et al., 2004), however previous computational modelling studies of reading acquisition demonstrate that differences in the orthographic transparency of a language can have implications for the distribution of labour between phonological, semantic and orthographic processing networks (Harm & Seidenberg, 2004; Yang et al., 2013). Without knowing the emergent consequences of the additional orthographic mapping performed by such networks, it is not possible to rule out the possibility that such training could result in a modulation of semantic competitor effects without requiring a reduction in cognitive efficiency to be implemented.

It is important to emphasise that the ASR model was not specifically designed to simulate effects of literacy, but rather was an effective model of multimodal effects on language processing that was coopted to extend to testing theories of literacy. However, a further means of validating the ASR model is to examine its additional predictions. Our results suggest that the representations governing language mediated eye gaze in low literates are more coarse grained and therefore gaze is less sensitive to the temporal location of phonological overlap. One means of testing the coarseness of low literates' phonological representations would be through examining their sensitivity to phonological rhyme overlap. It has been previously observed that literate individuals display greater sensitivity to phonological overlap in the onset of words than in the rhyme (Allopenna et al., 1998). This may be because, in the case of rhyme competitors, by the time later overlapping phonemes unfold, earlier phonological information can be used to eliminate the rhyme competitor as a possible target, hence onset competitors are fixated more than rhyme competitors. If the model's coarse grain representations equate to word level representations in low literates then such representations would be unsuitable for determining whether the overlap occurs in the onset or rhyme of a given word. If onset and rhyme are matched for length, then the word level representations will be equally similar and therefore will generate fixations equally in a network processing at this word level. The model therefore predicts that unlike high literates, low literates should display little difference in their bias towards fixating phonological onset and phonological rhyme competitors.

A second prediction of the effects of orthographic transparency on language mediated visual attention also follows from the framework outlined in this paper. Our simulations indicated that training on orthographic mapping is critical to developing more fine grained processing of phonological information and subsequently displaying increased sensitivity to phonological competitors. The processing level model on which our hypothesis was based, psycholinguistic grain size theory, posits that the level of transparency between orthography and phonology determines the granularity of processing that is developed. It then follows that in non-alphabetic languages, in which there is little componentiality in the correspondence between the orthography and the speech sounds that make up a word, literacy training will have little effect on the granularity of speech processing. Therefore, logographic literates should behave more like illiterates on tasks that aim to measure this aspect of processing. There is already substantial empirical evidence within the literature to support such a position with Chinese literates who have not been exposed to an alphabetic writing system displaying reduced levels of phonological awareness (Cheung et al., 2001; Ho & Bryant, 1997; Huang & Hanley, 1995, 1997; McBride-Chang et al., 2004; Read et al., 1986; Shu et al., 2008). Further, recent evidence from neuroimaging studies support the critical role of orthographic transparency in modulating effects of literacy on speech processing, with less involvement of associated orthographic processing regions observed in logographic literates compared to alphabetic literates when processing speech (Cao et al., 2011) and greater developmental changes in phonological processing regions as a consequence of literacy training in English over Chinese students (Brennan et al., 2013).

These results suggest that there may be similarities between illiterates and logographic literates in their phonological processing, and are compatible with the argument that literacy training on an alphabetic language leads to rearrangement in phonological processing networks such that phonological processing becomes more fine grained. Should the phonological effect observed in studies of language mediated eye gaze be modulated by the granularity of speech processing in the manner our simulations suggest, then we would predict that logographic literates (not exposed to training on alphabetic systems e.g. Hanyu Pinyin) like illiterates should display reduced sensitivity to phonological overlap compared to alphabetic literates, and, unlike alphabetic literates, their fixation of such competitors should not be tightly time locked to overlap in the speech signal. However if, as our simulations indicate, quantitative differences in the semantic bias observed between high and low literates result from increased efficiency of information transfer within the networks trained during literacy acquisition, we would not predict a reduction in semantic bias in logographic literate populations, as training of relevant networks should be similar in both literate groups.

To conclude, influential models of human language processing have been developed largely only with reference to the behaviour of alphabetic literates, and generally do not take into account the influence of literacy, or of varying orthographic systems, on the processing system. Those that currently do are likely to underrepresent its consequences because of their inclusion of pre-specified componential phonological representations. Our modelling work using the ASR model demonstrated that two competing theories of effects of literacy on language learning may well be compatible and complementary contributors to language processing: both cognitive efficiency and phonological grain-size differences were required to simulate the detailed data on phonological and semantic processing in literate and illiterate participants. Given that approximately 16% of the adult human population are illiterate (UNESCO Institute for Statistics, 2013) and a further 15%[2] (approximately) of the human population are literate in logographic languages, understanding the consequences of these factors for human cognition remains an important challenge for future research, for which multimodal computational models are likely to provide an informative, even necessary, tool.

References

Adrián, J. A., Alegria, J., & Morais, J. (1995). Metaphonological abilities of Spanish illiterate adults. *International Journal of Psychology, 30*(3), 329–351.

Alcock, K. J., Ngorosho, D., Deus, C., & Jukes, M. C. H. (2010). We don't have language at our house: Disentangling the relationship between phonological awareness, schooling, and literacy. *British Journal of Educational Psychology, 80*(1), 55–76.

Allopenna, P. D., Magnuson, J. S., & 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*(4), 419–439.

Anthony, J. L., & Francis, D. J. (2005). Development of phonological awareness. *Current Directions in Psychological Science, 14*(5), 255–259.

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language, 59*(4), 390–412.

Barr, D. J. (2008). Analyzing 'visual world' eyetracking data using multilevel logistic regression. *Journal of Memory and Language, 59*(4), 457–474.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language, 68*(3), 255–278.

Bengtsson, S. L., Nagy, Z., Skare, S., Forsman, L., Forssberg, H., & Ullén, F. (2005). Extensive piano practicing has regionally specific effects on white matter development. *Nature Neuroscience, 8*(9), 1148–1150.

Bramao, I., Mendonca, A., Faisca, L., Ingvar, M., Petersson, K. M., & Reis, A. (2007). The impact of reading and writing skills on a visuo-motor integration task: A comparison between illiterate and literate subjects. *Journal of the International Neuropsychological Society, 13*(2), 359–364.

Brennan, C., Cao, F., Pedroarena-Leal, N., McNorgan, C., & Booth, J. R. (2013). Reading acquisition reorganizes the phonological awareness network only in alphabetic writing systems. *Human Brain Mapping, 34*(12), 3354–3368.

Burnham, D. (2003). Language specific speech perception and the onset of reading. *Reading and Writing, 16*(6), 573–609.

Cao, F., Khalid, K., Lee, R., Brennan, C., Yang, Y., Li, K., et al (2011). Development of brain networks involved in spoken word processing of Mandarin Chinese. *NeuroImage, 57*(3), 750–759.

Caravolas, M., Volin, J., & Hulme, C. (2005). Phoneme awareness is a key component of alphabetic literacy skills in consistent and inconsistent orthographies: Evidence from Czech and English children. *Journal of Experimental Child Psychology, 92*, 107–139.

Chéreau, C., Gaskell, M. G., & Dumay, N. (2007). Reading spoken words: Orthographic effects in auditory priming. *Cognition, 102*(3), 341–360.

Cheung, H., Chen, H. C., Lai, C. Y., Wong, O. C., & Hills, M. (2001). The development of phonological awareness: Effects of spoken language experience and orthography. *Cognition, 81*(3), 227–241.

Cooper, R. M. (1974). The control of eye fixation by the meaning of spoken language: A new methodology for the real-time investigation of speech perception, memory, and language processing. *Cognitive Psychology, 6*(1), 84–107.

da Silva, C. G., Petersson, K. M., Faísca, L., Ingvar, M., & Reis, A. (2004). The effects of literacy and education on the quantitative and qualitative aspects of semantic verbal fluency. *Journal of Clinical and Experimental Neuropsychology, 26*(2), 266–277.

Dahan, D., & Tanenhaus, M. K. (2005). Looking at the rope when looking for the snake: Conceptually mediated eye movements during spoken-word recognition. *Psychonomic Bulletin & Review, 12*(3), 453–459.

de Jong, P. F., & van der Leij, A. (2003). Developmental changes in the manifestation of a phonological deficit in dyslexic children learning to read a regular orthography. *Journal of Educational Psychology, 95*(1), 22–40.

Deary, I. J., Bastin, M. E., Pattie, A., Clayden, J. D., Whalley, L. J., Starr, J. M., et al (2006). White matter integrity and cognition in childhood and old age. *Neurology, 66*(4), 505–512.

Dehaene, S., Pegado, F., Braga, L. W., Ventura, P., Nunes Filho, G., Jobert, A., et al (2010). How learning to read changes the cortical networks for vision and language. *Science, 330*(6009), 1359–1364.

Dijkstra, T., Roelofs, A., & Fieuws, S. (1995). Orthographic effects on phoneme monitoring. *Canadian Journal of Experimental Psychology, 49*(2), 264–271.

Engel, A. K., Fries, P., & Singer, W. (2001). Dynamic predictions: Oscillations and synchrony in top–down processing. *Nature Reviews Neuroscience, 2*(10), 704–716.

Fields, R. D. (2008). White matter in learning, cognition and psychiatric disorders. *Trends in Neurosciences, 31*(7), 361–370.

Garlock, V. M., Walley, A. C., & Metsala, J. L. (2001). Age-of-acquisition, word frequency, and neighborhood density effects on spoken word recognition by children and adults. *Journal of Memory and Language, 45*(3), 468–492.

[2] Based on a world population of 7.148 billion (United States Census Bureau, World Population Clock, 2014) and a population of the Peoples' Republic of China aged over 15 years 1.108 billion (The World Bank, 2012) of which 94% are literate (UNESCO Institite for Statistics., 2012).

Gaskell, M. G., & Marslen-Wilson, W. D. (1997). Integrating form and meaning: A distributed model of speech perception. *Language and Cognitive Processes, 12*(5–6), 613–656.
Goswami, U. (2003). Why theories about developmental dyslexia require developmental designs. *Trends in Cognitive Sciences, 7*(12), 534–540.
Gutiérrez, R., Boison, D., Heinemann, U., & Stoffel, W. (1995). Decompaction of CNS myelin leads to a reduction of the conduction velocity of action potentials in optic nerve. *Neuroscience Letters, 195*(2), 93–96.
Harm, M. W., & Seidenberg, M. S. (1999). Phonology, reading acquisition, and dyslexia: Insights from connectionist models. *Psychological Review, 106*(3), 491–528.
Harm, M. W., & Seidenberg, M. S. (2004). Computing the meanings of words in reading: Cooperative division of labor between visual and phonological processes. *Psychological Review, 111*(3), 662.
Ho, C. S. H., & Bryant, P. (1997). Phonological skills are important in learning to read Chinese. *Developmental Psychology, 33*(6), 946.
Hoonhorst, I., Medina, V., Colin, C., Markessis, E., Radeau, M., Deltenre, P., et al (2011). Categorical perception of voicing, colors and facial expressions: A developmental study. *Speech Communication, 53*(3), 417–430.
Huang, H. S., & Hanley, J. R. (1995). Phonological awareness and visual skills in learning to read Chinese and English. *Cognition, 54*(1), 73–98.
Huang, H. S., & Hanley, J. R. (1997). A longitudinal study of phonological awareness, visual skills, and Chinese reading acquisition among first-graders in Taiwan. *International Journal of Behavioral Development, 20*(2), 249–268.
Huettig, F., & Altmann, G. (2005). Word meaning and the control of eye fixation: Semantic competitor effects and the visual world paradigm. *Cognition, 96*(1), B23–B32.
Huettig, F., & Altmann, G. T. (2007). Visual-shape competition during language-mediated attention is based on lexical input and not modulated by contextual appropriateness. *Visual Cognition, 15*(8), 985–1018.
Huettig, F., & McQueen, J. M. (2007). The tug of war between phonological, semantic and shape information in language-mediated visual search. *Journal of Memory and Language, 57*(4), 460–482.
Huettig, F., Quinlan, P. T., McDonald, S. A., & Altmann, G. T. M. (2006). Models of high-dimensional semantic space predict language-mediated eye movements in the visual world. *Acta Psychologica, 121*(1), 65–80.
Huettig, F., Rommers, J., & Meyer, A. S. (2011). Using the visual world paradigm to study language processing: A review and critical evaluation. *Acta Psychologica, 137*, 151–171.
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*(6), 572–577.
Hulme, C., Snowling, M., Caravolas, M., & Carroll, J. (2005). Phonological skills are (probably) one cause of success in learning to read: A comment on Castles and Coltheart. *Scientific Studies of Reading, 9*(4), 351–365.
Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language, 59*(4), 434–446.
Kail, R., & Salthouse, T. A. (1994). Processing speed as a mental capacity. *Acta Psychologica, 86*(2), 199–225.
Kello, C. T., & Plaut, D. C. (2000). Strategic control in word reading: Evidence from speeded responding in the tempo-naming task. *Journal of Experimental Psychology. Learning, Memory, and Cognition, 26*(3), 719.
Kolinsky, R., Cary, L., & Morais, J. (1987). Awareness of words as phonological entities: The role of literacy. *Applied Psycholinguistics, 8*(3), 223–232.
Kolinsky, R., Pattamadilok, C., & Morais, J. (2012). The impact of orthographic knowledge on speech processing. *Ilha do Desterro A Journal of English Language, Literatures in English and Cultural Studies, 63*, 161–186.
Kosmidis, M. H., Tsapkini, K., Folia, V., Vlahou, C. H., & Kiosseoglou, G. (2004). Semantic and phonological processing in illiteracy. *Journal of the International Neuropsychological Society, 10*(6), 818–827.
Levi-Bruhl (1923). In *Primitive mentality. Translated from the French by Lilian A. Clare*. London: George Allen & Unwin. (Orig. pub. Paris, 1922).
Li, S. C., Lindenberger, U., Hommel, B., Aschersleben, G., Prinz, W., & Baltes, P. B. (2004). Transformations in the couplings among intellectual abilities and constituent cognitive processes across the life span. *Psychological Science, 15*(3), 155–163.
Li, Y., Liu, Y., Li, J., Qin, W., Li, K., Yu, C., et al (2009). Brain anatomical network and intelligence. *PLoS Computational Biology, 5*(5), e1000395.
Loureiro, C. D. S., Willadino Braga, L., Souza, L. D. N., Queiroz, E., & Dellatolas, G. (2004). Degree of illiteracy and phonological and metaphonological skills in unschooled adults. *Brain and Language, 89*(3), 499–502.
Luria, A. (1976). *Cognitive development: Its cultural and social foundations*. Cambridge: Cambridge University Press.
Madden, D. J., Bennett, I. J., & Song, A. W. (2009). Cerebral white matter integrity and cognitive aging: Contributions from diffusion tensor imaging. *Neuropsychology Review, 19*(4), 415–435.
Marslen-Wilson, W., & Tyler, L. K. (1980). The temporal structure of spoken language understanding. *Cognition, 8*(1), 1–71.
McBride-Chang, C., Bialystok, E., Chong, K. K., & Li, Y. (2004). Levels of phonological awareness in three cultures. *Journal of Experimental Child Psychology, 89*(2), 93–111.
McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology, 18*(1), 1–86.
Mirman, D., Dixon, J. A., & Magnuson, J. S. (2008). Statistical and computational models of the visual world paradigm: Growth curves and individual differences. *Journal of Memory and Language, 59*(4), 475–494.
Mirman, D., & Magnuson, J. S. (2009). Dynamics of activation of semantically similar concepts during spoken word recognition. *Memory & Cognition, 37*(7), 1026–1039.
Monaghan, P., & Shillcock, R. C. (2004). Hemispheric asymmetries in cognitive modelling: Connectionist modelling of unilateral visual neglect. *Psychological Review, 111*, 283–308.
Morais, J., Bertelson, P., Cary, L., & Alegria, J. (1986). Literacy training and speech segmentation. *Cognition, 24*(1), 45–64.
Morais, J., Cary, L., Alegria, J., & Bertelson, P. (1979). Does awareness of speech as a sequence of phones arise spontaneously? *Cognition, 7*(4), 323–331.

Morais, J., Content, A., Cary, L., Mehler, J., & Segui, J. (1989). Syllabic segmentation and literacy. *Language and Cognitive Processes, 4*(1), 57–67.

Morrison, F. J., Smith, L., & Dow-Ehrensberger, M. (1995). Education and cognitive development: A natural experiment. *Developmental Psychology, 31*(5), 789.

Muneaux, M., & Ziegler, J. (2004). Locus of orthographic effects in spoken word recognition: Novel insights from the neighbour generation task. *Language and Cognitive Processes, 19*(5), 641–660.

Norris, D., & McQueen, J. M. (2008). Shortlist B: A Bayesian model of continuous speech recognition. *Psychological Review, 115*(2), 357–395.

Norris, D., McQueen, J. M., & Cutler, A. (2000). Merging information in speech recognition: Feedback is never necessary. *Behavioral and Brain Sciences, 23*(03), 299–325.

Olivers, C. N. L., Huettig, F., Singh, J. P., & Mishra, R. K. (2014). The influence of literacy on visual search. *Visual Cognition, 21*, 74–101.

Pattamadilok, C., Knierim, I. N., Duncan, K. J. K., & Devlin, J. T. (2010). How does learning to read affect speech perception? *The Journal of Neuroscience, 30*(25), 8435–8444.

Pattamadilok, C., Perre, L., Dufau, S., & Ziegler, J. C. (2009). On-line orthographic influences on spoken language in a semantic task. *Journal of Cognitive Neuroscience, 21*(1), 169–179.

Penke, L., Maniega, S. M., Murray, C., Gow, A. J., Hernández, M. C. V., Clayden, J. D., et al (2010). A general factor of brain white matter integrity predicts information processing speed in healthy older people. *The Journal of Neuroscience, 30*(22), 7569–7574.

Perre, L., Midgley, K., & Ziegler, J. C. (2009a). When beef primes reef more than leaf: Orthographic information affects phonological priming in spoken word recognition. *Psychophysiology, 46*(4), 739–746.

Perre, L., Pattamadilok, C., Montant, M., & Ziegler, J. C. (2009b). Orthographic effects in spoken language: On-line activation or phonological restructuring? *Brain Research, 1275*, 73–80.

Perre, L., & Ziegler, J. C. (2008). On-line activation of orthography in spoken word recognition. *Brain Research, 1188*, 132–138.

Pujol, J., Soriano-Mas, C., Ortiz, H., Sebastian-Galles, N., Losilla, J. M., & Deus, J. (2006). Myelination of language-related areas in the developing brain. *Neurology, 66*(3), 339–343.

Read, C., Yun-Fei, Z., Hong-Yin, N., & Bao-Qing, D. (1986). The ability to manipulate speech sounds depends on knowing alphabetic writing. *Cognition, 24*(1), 31–44.

Reis, A., & Castro-Caldas, A. (1997). Illiteracy: A cause for biased cognitive development. *Journal of the International Neuropsychological Society, 3*(05), 444–450.

Reis, A., Guerreiro, M., & Petersson, K. M. (2003). A sociodemographic and neuropsychological characterization of an illiterate population. *Applied Neuropsychology, 10*(4), 191–204.

Salthouse, T. A. (1988). Initiating the formalization of theories of cognitive aging. *Psychology and Aging, 3*(1), 3–16.

Salthouse, T. A. (1996). The processing-speed theory of adult age differences in cognition. *Psychological Review, 103*(3), 403–428.

Salthouse, T. A. (2005). Relations between cognitive abilities and measures of executive functioning. *Neuropsychology, 19*(4), 532–545.

Scliar-Cabral, L., Morais, J., Nepomuceno, L., & Kolinsky, R. (1997). The awareness of phonemes: So close-so far away. *International Journal of Psycholinguistics, 13*, 211–240.

Serniclaes, W., Ventura, P., Morais, J., & Kolinsky, R. (2005). Categorical perception of speech sounds in illiterate adults. *Cognition, 98*(2), B35–B44.

Shu, H., Peng, H., & McBride-Chang, C. (2008). Phonological awareness in young Chinese children. *Developmental Science, 11*(1), 171–181.

Smith, A. C., Monaghan, P., & Huettig, F. (2013a). An amodal shared resource model of language-mediated visual attention. *Frontiers in Psychology, 4*.

Smith, A. C., Monaghan, P., & Huettig, F. (2013b). Modelling language-vision interactions in the hub-and-spoke framework. In J. Mayor & P. Gomez (Eds.), *Computational Models of Cognitive Processes: Proceedings of the 13th Neural Computation and Psychology Workshop (NCPW13)*. Singapore: World Scientific Publishing.

Stoodley, C. J., & Stein, J. F. (2006). A processing speed deficit in dyslexic adults? Evidence from a peg-moving task. *Neuroscience Letters, 399*(3), 264–267.

Storkel, H. L. (2002). Restructuring of similarity neighbourhoods in the developing mental lexicon. *Journal of Child Language, 29*(2), 251–274.

Szwed, M., Ventura, P., Querido, L., Cohen, L., & Dehaene, S. (2012). Reading acquisition enhances an early visual process of contour integration. *Developmental Science, 15*(1), 139–149.

Taft, M. (2006). Orthographically influenced abstract phonological representation: Evidence from non-rhotic speakers. *Journal of Psycholinguistic Research, 35*(1), 67–78.

Taft, M., & Hambly, G. (1985). The influence of orthography on phonological representations in the lexicon. *Journal of Memory and Language, 24*(3), 320–335.

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*(5217), 1632–1634.

The World Bank (2014, February 21). Development Indicators. Retrieved February, 21, 2014, from <http://data.worldbank.org/indicator/SP.POP.0014.TO.ZS/countries>.

Tolhurst, D. J., & Lewis, P. R. (1992). Effect of myelination on the conduction velocity of optic nerve fibres. *Ophthalmic and Physiological Optics, 12*(2), 241–243.

Treiman, R., & Zukowski, A. (1991). Levels of phonological awareness. In S. A. Brady & D. P. Shankweiler (Eds.), *Phonological processes in literacy: A tribute to Isabelle Y. Liberman* (pp. 67–83). Routledge: Oxford.

Turken, A. U., Whitfield-Gabrieli, S., Bammer, R., Baldo, J. V., Dronkers, N. F., & Gabrieli, J. D. (2008). Cognitive processing speed and the structure of white matter pathways: Convergent evidence from normal variation and lesion studies. *Neuroimage, 42*(2), 1032–1044.

UNESCO Institite for Statistics. (2012). Adult and Youth Literacy, 1990–2015, Montreal: UNESCO.

UNESCO Institute for Statistics. (2013). Adult and Youth Literacy Fact Sheet, Montreal: UNESCO.

United States Census Bureau. (2014, February 21). World Population Clock. Retrieved February, 21, 2014, from <http://www.census.gov/popclock/>.
Ventura, P., Kolinsky, R., Querido, J. L., Fernandes, S., & Morais, J. (2007). Is phonological encoding in naming influenced by literacy? *Journal of Psycholinguistic Research, 36*(5), 341–360.
Ventura, P., Morais, J., Pattamadilok, C., & Kolinsky, R. (2004). The locus of the orthographic consistency effect in auditory word recognition. *Language and Cognitive Processes, 19*(1), 57–95.
Vygotsky, L. S. (1978). Mind in society. In M. Cole, V. John-Steiner, S. Scribner, & E. Souberman (Eds.), *The development of higher psychological processes society*. Cambridge: Harvard University Press.
Waxman, S. G. (1980). Determinants of conduction velocity in myelinated nerve fibers. *Muscle and Nerve, 3*(2), 141–150.
Yang, J., McCandliss, B. D., Shu, H., & Zevin, J. D. (2009). Simulating language-specific and language-general effects in a statistical learning model of Chinese reading. *Journal of Memory and Language, 61*(2), 238–257.
Yang, J., Shu, H., McCandliss, B. D., & Zevin, J. D. (2013). Orthographic influences on division of labor in learning to read Chinese and English: Insights from computational modeling. *Bilingualism: Language and Cognition, 16*(2), 354–366.
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), 1–14.
Yu, C., & Ballard, D. H. (2007). A unified model of early word learning: Integrating statistical and social cues. *Neurocomputing, 70*, 2149–2165.
Ziegler, J. C., Bertrand, D., Tóth, D., Csépe, V., Reis, A., Faísca, L., et al (2010). Orthographic depth and its impact on universal predictors of reading a cross-language investigation. *Psychological Science, 21*, 551–559.
Ziegler, J. C., & Ferrand, L. (1998). Orthography shapes the perception of speech: The consistency effect in auditory word recognition. *Psychonomic Bulletin & Review, 5*(4), 683–689.
Ziegler, J. C., & Goswami, U. (2005). Reading acquisition, developmental dyslexia, and skilled reading across languages: A psycholinguistic grain size theory. *Psychological Bulletin, 131*(1), 3–29.