Modelling Language – Vision Interactions in the Hub and Spoke Framework

Alastair C. Smith
Max Planck Institute for Psycholinguistics,
Wundtlaan 1, 6525 XD Nijmegen, The Netherlands
**E-mail: alastair.smith@mpi.nl*

Padraic Monaghan
Department of Psychology,
Lancaster University, Lancaster LA1 4YF, UK

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

Multimodal integration is a central characteristic of human cognition. However our understanding of the interaction between modalities and its influence on behaviour is still in its infancy. This paper examines the value of the Hub & Spoke framework (Plaut, 2002; Rogers et al., 2004; Dilkina et al., 2008; 2010) as a tool for exploring multimodal interaction in cognition. We present a Hub and Spoke model of language–vision information interaction and report the model's ability to replicate a range of phonological, visual and semantic similarity word-level effects reported in the Visual World Paradigm (Cooper, 1974; Tanenhaus et al, 1995). The model provides an explicit connection between the percepts of language and the distribution of eye gaze and demonstrates the scope of the Hub-and-Spoke architectural framework by modelling new aspects of multimodal cognition.

1.1 Introduction

Hub and Spoke (H&S) models (Plaut, 2002; Rogers et al., 2004; Dilkina et al., 2008; 2010) are characterised by a central resource that integrates modality specific information. The approach reflects the increased interest and awareness within cognitive science of multimodal cognitive interactions. To date computational implementations of the H&S framework have been used in conjunction with neuropsychological data to offer both explanation for a range of semantic related neuropsychological disorders and insight into how semantic processing may be implemented within the brain. This paper aims to highlight the potential for broader application of this framework. Research within cognitive neuroscience has demonstrated the difficulty of assigning modality specific functions to distinct neural processing regions (see Anderson 2010; Poldrack, 2006). An increased understanding of how modality specific information may be integrated with information from other modalities and how such a system may behave could therefore prove valuable to neuropsychology. The H&S computational modelling framework offers a tool for investigating such complex interactive aspects of multimodal cognition.

1.2 Virtues of the Hub & Spoke Framework

When hearing a spoken word such as “apple” it is possible to bring to mind its visual form. When seeing an object such as an ‘apple’ it is also possible to bring to mind the spoken word used to describe the object “apple”. How are modality specific representations connected across modalities, what is the nature of representation in each modality and how are connections between representations acquired? Previous H&S models have offered answers to each of these questions.

The H&S framework has proved successful by providing a parsimonious architecture in which single modality models can be drawn together to examine the consequences of multimodal interaction and representation. Due to the complexity inherent in multimodal processing, predicting the connectivity between modalities without explicit

implementation can be challenging. For instance, an apparent dissociation between lexical and semantic performance in semantic dementia patients suggested the need for separate systems supporting lexical and semantic processing. However, the H&S offered a means of testing the compatibility of a fully integrated model with the behavioural data, and Dilkina et al. (2008, 2010) demonstrated that counter to previous assumptions the pattern of behaviour observed was consistent with a single system H&S model.

The H&S framework offers a single system architecture with minimal architectural assumptions, and this makes it possible to isolate the influence of two further major determinants of emergent behaviour in such complex multimodal systems, 1) the structure of representations and/or 2) the tasks or mappings demanded by the learning environment.

Plaut (2002) and Rogers, et al. (2004) present two alternative means of exploring the role of representational structure through use of the H&S framework. Plaut (2002) focused on a single aspect of representational structure, specifically systematic or arbitrary relationships between modalities. By abstracting out additional complexity within representations Plaut was able to investigate the emergent properties of this single factor. In contrast Rogers et al. (2004) and Dilkina et al. (2008, 2010) provided a richer representation of the structure available within the learning environment by deriving semantic representations from attribute norming studies. This enabled the authors to examine the emergent properties a single system multimodal model is capable of developing with such richer input. It is through simulating such complexity within the learning environment that their model was able to replicate the broad variability displayed by semantic dementia patients that had previously been viewed as challenging for single system accounts. Such approaches demonstrate the framework's potential for providing a more detailed understanding of how representational structure shapes multimodal cognition.

As the H&S framework allows a model to perform multiple mappings, decisions are required as to which mappings are performed, how frequently they are performed and how these variables might change over the course of development. Dilkina et al. (2010) introduced stages of development within the model training process. They attempted to

provide a more accurate depiction of the constraints placed on systems during development by manipulating the frequency and period in which given tasks are performed by the model (e.g., mapping orthography to phonology was only performed during the second stage of development). This is an example of how the framework can be used to explore the relationship between environmental constraints, such as the type and frequency of mappings performed during development, and the emergent behaviour displayed by the system.

To date the H&S framework has been used primarily in conjunction with neuropsychological data. This approach provides clear advantages when aiming to map network architecture onto neural populations and has brought significant progress in this direction with evidence emerging for a mapping of the semantic hub (integrative layer) onto neural populations in the anterior temporal lobe (Lambon-Ralph et al., 2010). The framework however also offers scope for examining the factors underlying individual differences within non-patient populations, a feature yet to be exploited. For example, as we have described, the framework makes it possible to examine how contrasts in the learning environment, be it in the input to the system (e.g., richness or diversity of input) or the mappings demanded (e.g., learning to read: orthography to phonology), can result in variation in behaviour both across development and in mature systems.

Further as multimodal integration is central to many aspects of human cognition the H&S framework has the potential to provide insight into many new areas of cognitive processing. The following section provides an example, utilizing the framework to model the influence of language-vision interactions on eye gaze behaviour.

1.3 A Hub & Spoke Model of Language Mediated Visual Attention

1.3.1 *Language Mediated Visual Attention & The Visual World Paradigm*

Within daily communication we experience the cognitive system's ability to rapidly and automatically integrate information from linguistic and non-linguistic streams. For example many spoken utterances become ambiguous without the context in which they are delivered (e.g., "I love this field!"). Many have placed profound importance on the role of multimodal interaction during language processing (Mayberry, Crocker & Knoeferle, 2009), arguing that it is such multimodal context that grounds communication in the world that we share.

The Visual World Paradigm (VWP) offers a means of examining online language-vision interactions and has led to a substantial literature examining the role of non-linguistic information in language processing (for review see Huettig et al., 2011). Within the paradigm, participants are exposed to a visual display and an auditory stimulus while their eye movements are recorded. Its application has led to a detailed description of how manipulation of relationships between visual and auditory stimuli can drive systematic variation in eye gaze behaviour.

Allopenna et al. (1998) demonstrated that the probability of fixating items within the visual display in a VWP could be modulated by phonological overlap between the name corresponding to the item depicted and the target word in the auditory stimulus. Given a target word (e.g. beaker), cohort competitors (e.g. beetle) were fixated earlier and with higher probability than rhyme competitors (e.g. speaker), both of which were fixated with higher probability than unrelated distractors (e.g. carriage). Simulations using the TRACE model of speech perception (McClelland & Elman, 1986) replicated both the time course and probability of fixations displayed by participants.

Visual relationships between spoken words and displayed items have also been shown to influence fixation behaviour in the VWP. Dahan and Tanenhaus (2005) and Huettig and Altmann (2007) presented within a visual display items that shared visual features (e.g. snake) with target

items (e.g. rope). They observed that participants fixated such visual competitors with higher probability than unrelated distractors both when the target was present in the display (Dahan & Tanenhaus, 2005) and when the target was absent (Huettig & Altmann, 2007).

Semantics provides a third dimension in which overlap between target and competitor has been shown to modulate fixation behaviour. Huettig and Altmann (2005) and Yee and Sedivy (2006) both demonstrated that items within the visual display that shared overlapping sematic features with a target word were fixated with higher probability than unrelated distractors. Again this was demonstrated in conditions in which the target was present (Yee & Sedivy, 2006) and when the target was absent (Huettig and Altmann, 2005).

Taken together this evidence indicates that language mediated eye gaze can be driven by representational overlap in visual, semantic and phonological dimensions. Although such studies provide a detailed description of how eye gaze varies as a function of the relationships between visual and auditory stimuli we know very little about the processes connecting these events. Issues such as what is the nature of representation in each modality, how are representations across modalities connected and how is activation of such representations connected to eye gaze behaviour remain unresolved.

The behaviour of participants in the VWP is dependent on language-vision interactions, the interaction of visual information from the visual display, and auditory information from a spoken utterance. Although models of language vision interaction applied to the VWP (Spivey, 2008; Mayberry et al, 2009; Kukona & Tabor, 2010) exist, they lack sufficient depth of representation in visual, semantic and phonological dimensions to provide explanation for the range of word level effects observed in the VWP. Similarly single modality models of VWP (Allopenna et al, 1998; Mirman & Magnuson, 2009) although providing depth of representation in a single modality, lack the necessary representation in other modalities that is required to offer a comprehensive description of language mediated eye gaze.

Our current study aims to explore the scope of the H&S framework as a tool for examining multimodal cognition by using it to derive a model of language mediated eye gaze that provides an explicit connection

between the percepts of language and the distribution of eye gaze. We constructed a recurrent neural network that integrates semantic, visual and phonological information within a central resource and test its ability to capture the range of word level effects observed in the VWP, described in this section.

1.3.2 *Method*

1.3.2.1 *Network*

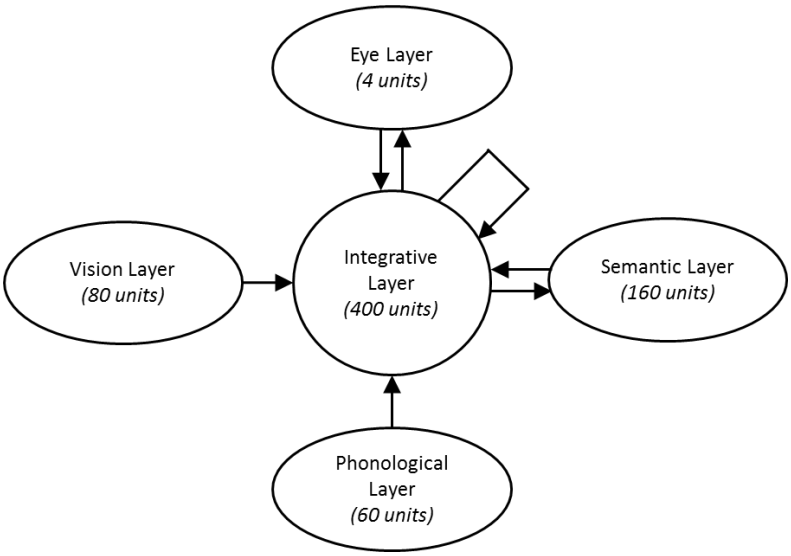

Figure 1. Network Architecture

The architecture of the neural network used in this study is displayed in Figure 1. It consists of a central resource that integrates modality specific information from four visible layers. The phonological layer is composed of six phoneme slots, each represented by 10 units and models the input of phonological information extracted from the speech signal. The vision layer encodes visual information extracted from four locations in the visual display each location represented by 20 units. The central

integrative layer is self-connected and has both feed-forward and feed-back connections to both semantic and eye layers. The semantic layer consists of 160 units and allows the model to capture the semantic properties of items. The eye layer is defined by 4 units, with each unit associated with a given location in the visual field. Activation of each eye unit is taken to represent the probability of fixating the associated spatial location in the visual display. Layers are fully connected in all specified directions shown in Figure 1.

1.3.2.2 *Artificial Corpus*

The artificial corpus was constructed taking a fundamentalist approach (see Plaut, 2002) to allow control over relationships between representations embedded within and across modalities. The corpus consisted of 200 items each with unique phonological, visual and semantic representations.

Phonological representations were produced by pseudo-randomly assigning a fixed sequence of six phonemes to each word with each phoneme taken from a phoneme inventory consisting of 20 possible phonemes. Each phoneme was defined by a 10 unit binary vector. As was the case with semantic and visual representations binary values were assigned with p = 0.5. Visual representations were defined by 20 unit binary vectors. Semantic representations were encoded by 160 unit binary vectors. Semantic representations were sparsely distributed, with each item assigned 8 of a possible 160 semantic properties.

Table 1. Controls on signal overlap within artificial corpus

Modality	**Item**	**Constraint**	**Signal Overlap ($\bar{x}$)**
Phonology	Competitor	3 of 6 phonemes shared with target	75%
	Unrelated	Max. 2 consecutive phonemes shared with any other item	50%
Function	Near	4 of 8 functional properties shared with target	97.5%
	Unrelated	Max. 1 functional property shared with target	95%
Vision	Competitor	For 10 visual features P(feature overlap with target) = 1; for remaining features P(feature overlap with target) = 0.5	75%
	Unrelated	P(feature overlap with target) = 0.5	50%

Details of the constraints on the construction of representations applied within each modality can be found in Table 1. Onset competitors shared their initial sequence of three phonemes with target items, while rhyme competitors shared the final sequence of three phonemes with targets. All other phonological representations overlapped by a maximum of two consecutive phonemes. Visual competitors shared 10 visual features with targets and overlapped with p = 0.5 across the remaining 10 features. Semantic neighbours shared 4 of 8 semantic properties with target items.

1.3.2.3 *Training*

The model was trained on four tasks chosen to simulate the tasks performed by participants prior to testing through which associations between representations are acquired. The model was trained to map from visual representations to semantic representations, from phonological representations to semantic representations and to activate the eye unit corresponding to the location of the visual representation of a target defined by the presence of its phonological or semantic form. Details of the timing of events within each training task can be found in Table 2. In tasks that involved the presentation of phonological information phonemes were presented to the model sequentially, with one additional phoneme presented at each subsequent time point.

Training tasks were assigned trial by trial on a pseudo random basis. To reflect our assumption that individuals select items based on their associated semantic properties more frequently than selecting items based on an external auditory stimulus, phonology driven orienting tasks occurred four times less than all other training tasks. Items were selected from the corpus and assigned locations and/or roles (target/distractor) randomly. All connection weights were initially randomised and adjusted during training using recurrent back-propagation (learning rate = 0.05). Training was terminated after 1 million trials.

Table 2. Temporal organisation of events in model training

Task	Vision (Vis)		Phonology (Phon)		Semantics (Sem)		Eye	
	Description	**ts**	**Description**	**ts**	**Description**	**ts**	**Description**	**ts**
Vis to Sem	4 visual representations randomly selected from corpus, 1 assigned as target	0-14	Random time invariant noise provided as input	0-14	*Semantic representation of target provided post display onset*	*3-14*	Location of target activated, all other locations inactive	0-14
Phon to Sem	Random time invariant noise provided as input across all 4 input slots	0–14	Phonology of target provided as a staggered input	0-14	*Semantic representation of target provided post phonological disambiguation*	*5-14*	No constraints on activation	-
Phon to Location	4 visual representations randomly selected from corpus, 1 assigned as target	0-14	Phonology of target provided as a staggered input	0-14	No constraints on activation	-	*Post disambiguation location of target activated, all other locations inactive*	*5-14*
Sem to Location	4 visual representations randomly selected from corpus, 1 assigned as target	0-14	Random time invariant noise provided as input	0-14	Semantic representation of target provided	0-14	*Location of target activated, all other locations inactive post functional onset*	*2-14*

1.3.2.4 *Pre-Test*

Results presented in this paper reflect the mean performance of six instantiations of the model. Once trained the model was tested on its ability to perform each of the four training tasks for all items in all locations. In mapping from visual or phonological form to semantics, activation of the semantic layer was most similar (cosine similarity) to that of the target for 100% of items. For both phonological and semantic orientation tasks the eye unit associated with the location corresponding to that of the target was most active for 98% of items.

1.3.3 ***Results***

Within the following simulations input to the visual layer was provided at time step (ts) 0 and remained until the end of each trial (ts 29). Figures display the average activation of the eye unit corresponding to the location of the given competitor type as a proportion of the total activation of all units within the eye layer. Onset of the target phonology

occurs at ts 5, with a single additional phoneme provided at each subsequent ts. For each simulation 20 experimental sets were constructed, with each set containing 4 items. The model was then tested on all 20 sets with each set tested in all 24 possible arrangements of item and location. Comparisons of the probability of fixating each competitor type were calculated using the procedure described in Huettig & McQueen (2007). The mean probability of fixating a given type of competitor was calculated across all time points and across all trials in a given set. The ratio between the mean probability of fixating a given competitor type and the sum of the probability of fixating the given competitor and the competitor against which it is to be compared was calculated. This ratio was then compared across sets to 0.5 using a one-sample t-test.

1.3.3.1 *Simulation of Phonological Effects*

To simulate Allopenna et al., (1998) the model was presented with a display containing the visual representation of a target, an onset competitor (first three phonemes overlap with target), a rhyme competitor (final three phonemes overlap with target) and an unrelated distractor (no phonological overlap with target) (see Figure 2A). All items were controlled for semantic and visual similarity. Both onset competitors [mean ratio = 0.57, $t(19) = 9.86$, $p < 0.001$] and targets [mean ratio = 0.76, $t(19) = 54.84$, $p < 0.001$] were fixated with higher probability than unrelated distractors, although onset competitors were fixated less than the targets [mean ratio = 0.30, $t(19) = -56.91$, $p < 0.001$]. These patterns of fixation behaviour replicate those reported in Allopenna et al., (1998). However in contrast to their findings the probability of fixating rhyme competitors did not differ from that of unrelated distractors.

Although such conditions were not tested in Allopenna et al., (1998) simulations were also run in which the target item was replaced by an additional unrelated distractor (Figure 2B). As in the target present condition onset competitors were fixated more than distractors [mean ratio = 0.57, $t_2(19) = 13.11$, $p < 0.001$], while the probability of fixating rhyme competitors and unrelated distractors did not differ.

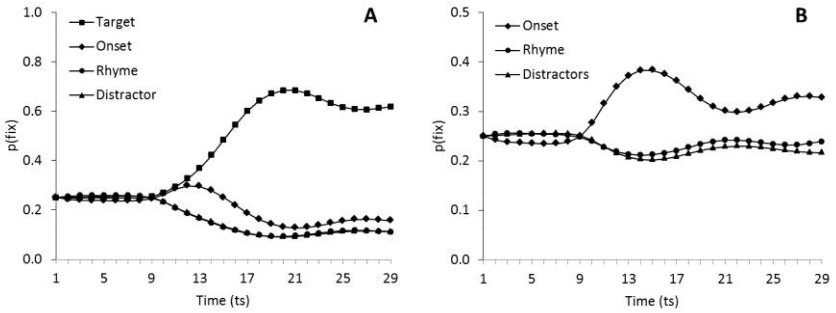

Figure 2. Mean proportional activation of eye units across all items and locations A) phonological competitors & target item; B) phonological competitors only (target absent)

1.3.3.2 *Simulation of Visual Effects*

In Dahan & Tanenhaus (2005) participants are presented with displays containing a target, a visual competitor and two unrelated distractors, these conditions were simulated within the model with results presented in Figure 3A. The model replicates the behaviour of participants reported in Dahan & Tanenhaus (2005) with both target items [mean ratio = 0.76, $t(19) = 82.70$, $p < 0.001$] and visual competitors [mean ratio = 0.61, $t(19) = 23.97$, $p < 0.001$] fixated with higher probability than unrelated distractors. The model also fixated visual competitors [mean ratio = 0.33, $t(19) = -47.19$, $p < 0.001$] less than target items again replicating the behaviour of participants.

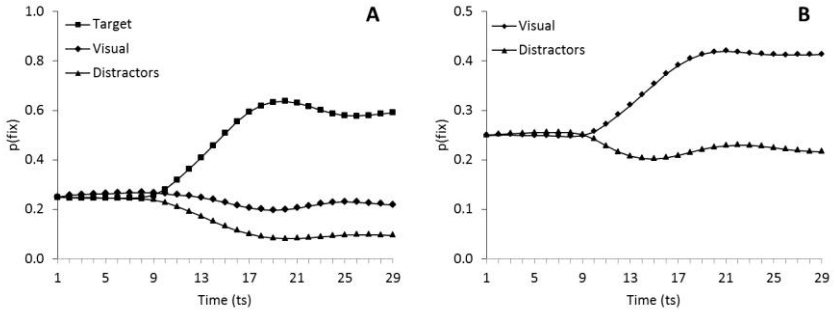

Figure 3. Mean proportional activation of eye units across all items and locations A) visual competitor & target item; B) visual competitor only (target absent)

Huettig & Altmann (2007) tested participants using displays containing a single visual competitor in addition to three unrelated distractors, this was also simulated in the model with results shown in Figure 3B. Again the model successfully replicates the behaviour displayed by participants, fixating visual competitors with higher probability than unrelated items [mean ratio = 0.60, $t(19) = 24.43$, $p < 0.001$].

1.3.3.3 *Simulation of Semantic Effects*

Simulations were also run of conditions reported in Yee and Sedivy (2006) and Huettig and Altmann (2005) in which participants are presented with scenes containing a target, a semantic competitor and two unrelated distractors (see Figure 4A). The model replicated participants increased probability of fixating target items [mean ratio = 0.76, $t(19) = 96.44$, $p < 0.001$] and semantic competitors [mean ratio = 0.58, $t(19) = 12.73$, $p < 0.001$] in comparison to unrelated distractors. Also as with participants semantic competitors were fixated less that target items [mean ratio = 0.30, $t(19) = -50.04$, $p < 0.001$].

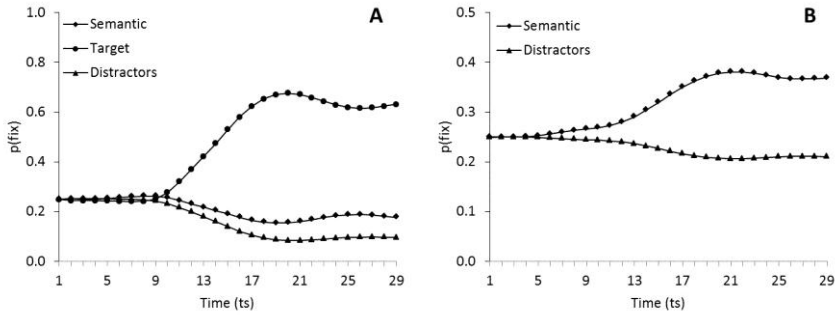

Figure.4. Mean proportional activation of eye units across all items and locations A) semantic competitor & target item; B) semantic competitor only (target absent)

Huettig & Altmann (2005) also report the participants' fixation behaviour when presented with scenes in which the target is replaced by an additional unrelated distractor. Again the model replicates the pattern of increased fixation towards sematic competitors over unrelated items

[mean ratio = 0.58, t(19) = 14.46, p <0.001] reported by Huettig & Altmann (2005).

1.4 Discussion

The results demonstrated that the model of language – vision interactions proposed within this paper is capable of replicating a broad range of single modality, word-level effects reported within the VWP literature. The model successfully replicates phonological onset, visual and semantic competitor effects in both target present and target absent conditions.

Although it does not precisely simulate the rhyme effects observed in Allopenna et al. (1998), we do not believe that this constitutes a significant problem for the proposed framework. Rhyme effects reported within the literature tend to be weak and far less robust than the onset effects the model does successfully replicate (see Allopenna et al., 1998; Desroches et al., 2006; McQueen & Viebahn, 2007; McQueen & Huettig, 2012). Moreover McQueen & Huettig (2012) provides evidence that the comparative onset rhyme effect is modulated by the level of intermittent noise present in the speech signal. It is therefore possible such subtle effects are beyond the scope of the current model. However we do offer an alternative explanation in which the current framework remains compatible with the observed rhyme effects. In the current model input is a perfect replication of the corresponding auditory or visual representation. Therefore, initial phonological input always corresponds to the target item. However for real world stimuli the perceptual system is frequently provided with impoverished or noisy representations. We believe that training the model on a more realistic representation of the stimuli available in the true learning environment may lead to the emergence of rhyme effects. A model trained on noisy auditory input would rely less wholly on early sections of the speech signal, for sections of this signal may not correspond to the target representation. Consequently in comparison to the current model, greater value would be placed on later sections of the signal as they may contain additional valuable information that would allow the model to identify

the target. Future work could explore this aspect of the learning environment through the addition of noise to model input; we predict that its introduction would reduce the influence on attention of initial phonemes and make it more likely for attention to be drawn by items with consistent overlap in later phonemes.

The above study provides an example of how single modality models can be integrated through the H&S framework to explore multimodal cognitive interactions. Although the connectivity between modalities is likely to be far more complex than that captured by the current model (McNorgan, Reid & McRae, 2011), our results indicate that the current framework can act as a good initial proxy for this interaction.

Given the models success in replicating the VWP data this would suggest that the model successfully captures some true aspect of the underlying cognitive processes recruited during completion of such tasks. The model presented offers an explicit description of the connection between the percepts of language and the distribution of eye gaze. This includes a description of the structure of representation within each modality, the features of those representations that drive competitor effects, the mechanisms that connect representations across modalities and a description of how such connections emerge from the structure of the learning environment. VWP research has moved on to examining multimodal effects in language-vision interactions (see Huettig & McQueen, 2007). We hope to extend the current study by exploring the extent to which hypotheses implemented within the above model also offer explanation for such multimodal effects.

A further stage in this research project will be to exploit additional virtues of the H&S framework described in earlier sections of this paper to investigate the mechanisms that underlie individual differences in language mediated eye gaze. Recent evidence suggests that environmental factors such as exposure to formal literacy training can have a significant effect on language mediated visual attention (Huettig, Singh & Mishra, 2011). Through use of the current model we aim to examine which environmental factors could give rise to the variation in behaviour observed and provide an explicit description of how such factors drive variation in eye gaze behaviour.

References:

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, pp 419-439.

Anderson, M. L. (2010). Neural reuse: A fundamental organizational principle of the brain. *Behavioural and Brain Sciences*, 33(4), pp 245-266.

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, pp 84−107.

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, pp 453−459.

Desroches, A. S., Joanisse, M. F., and Robertson, E. K. (2006). Specific phonological impairments in dyslexia revealed by eyetracking. *Cognition*, 100, B32–B42.

Dilkina, K., McClelland, J.L., Plaut, D.C., (2008). A single-system account of semantic and lexical deficits in five semantic dementia patients. *Cognitive. Neuropsychology*. 25, pp 136–164.

Dilkina, K., McClelland, J. L., and Plaut, D. C. (2010). Are there mental lexicons? The role of semantics in lexical decision. *Brain Research*. 1365, pp 66-81.

Huettig, F., Altmann, G. T. M., (2007). Visual-shape competition during language mediated attention is based on lexical input and not modulated by contextual appropriateness. *Visual Cognition*, 15, pp 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, pp 460-482.

Huettig, F., Altmann, G. T. M., (2005). Word meaning and the control of eye fixation: Semantic competitor effects and the visual world paradigm. *Cognition*, 96, B23−B32.

Huettig, F., Rommers, J., Meyer, A., (2011). Using the visual world paradigm to study language processing: a review and critical evaluation. *Acta Psychologica*, 137, pp 151-171.

Huettig, F., Singh, N., Mishra, R. K., (2011). Language-mediated visual orienting behaviour in low and high literates. *Frontiers in Psychology*, 2, 285.

Kukona, A., & Tabor, W. (2011). Impulse processing: A dynamical systems model of incremental eye movements in the visual world paradigm. *Cognitive Science*, 35(6), 1009-1051.

Lambon-Ralph, M. A., Sage, K., Jones, R. W., and Mayberry, E. J. (2010). Coherent concepts are computed in the anterior temporal lobes. *PNAS*, 107 (6), pp 2717-2722.

Mayberry, M. R., Crocker, M. W., & Knoeferle, P. (2009). Learning to attend: A connectionist model of situated language comprehension. *Cognitive science*, 33(3), 449-496.

McClelland, J. L., Elman, J. L., (1986). The TRACE model of speech perception. *Cognitive Psychology*, 18, pp 1-86.

McNorgan, C., Reid, J., & McRae, K. (2011). Integrating conceptual knowledge within and across representational modalities. *Cognition*, 118, pp 211- 233.

McQueen, J. M., & Huettig, F. (2012). Changing only the probability that spoken words will be distorted changes how they are recognized. *Journal of the Acoustical Society of America*, 131(1), pp 509-517.

McQueen, J. M., and Viebahn, M. (2007). Tracking recognition of spoken words by tracking looks to printed words. *Quarterly Journal of Experimental Psychology*. 60, pp 661–671.

Mirman, D., & Magnuson, J. S. (2009). Dynamics of activation of semantically similar concepts during spoken word recognition. *Memory & cognition*, 37(7), 1026-1039.

Plaut, D. C. (2002). Graded modality-specific specialization in semantics: A computational account of optic aphasia. *Cognitive Neuropsychology*, 19, pp 603-639.

Poldrack, R. A. (2006) Can cognitive processes be inferred from neuroimaging data? *Trends in Cognitive Sciences*, 10, pp 59–63.

Rogers, T.T., Lambon Ralph, M.A., Garrard, P., Bozeat, S., McClelland, J.L., Hodges, J.R., & Patterson, K. (2004). The structure and deterioration of semantic memory: a neuropsychological and computational investigation. *Psychological Review*, 111, pp 205–235.

Spivey, M. (2008). The continuity of mind (Vol. 40). Oxford University Press, USA.

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, pp 1632−1634.

Yee, E., & Sedivy, J. C. (2006). Eye movements to pictures reveal transient semantic activation during spoken word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 32, pp 1−14.