Computational Modeling of Eye Movements – from Reading to Scene Viewing

Hsueh-Cheng Wang

Department of Computer Science, University of Massachusetts Boston

1. Introduction

In our everyday life, visual information is essential for our interaction with the environment, and sometimes even for our survival. For example, we *read* the newspaper, books, and web pages for retrieving, learning, and comprehending information and ideas. We continually shift our gaze to *inspect scenes* to understand the real world, or search for an object, e.g., look for a key. We *pay attention* to traffic signs or displays showing directions to a hospital or grocery store. Each task involves very complex processes in our visual system as well as in higher cognitive functions, and these processes rarely occur isolated. Scientists in different fields, such as artificial intelligence, linguistics, and psychophysics, are devoted to a common goal: to understand the nature of human vision and its relation to cognition and mind. Therefore, it is important to bring together investigations from diverse disciplines and perspectives.

My dissertation focuses on developing computational models of eye movements for understanding how cognitive processes (e.g., visual information processing, word recognition, attention, and oculomotor control) can work together to perform a complex everyday task (e.g., reading or scene viewing). In this theoretical framework, many biologically-inspired computational methods were used and found psychologically plausible to predict human behaviors and simulate human cognition.

For eye movements in reading, I proposed models of visual encoding, word identification, and semantic integration in contexts. In Chapter 2, singular value decomposition (SVD; Strang, 1993) was used to predict the most important strokes for Chinese character recognition (Wang, Angele, Schotter, Yang, Simovici, Pomplun, & Rayner, under minor revision). In Chapter 3, I used latent semantic analysis (LSA, see Landauer & Dumais, 1997) to explain how readers rate the semantic transparency of English and Chinese compound words. A linear regression model was developed to estimate contextual predictability during reading (Wang, Pomplun, Chen, Ko, & Rayner, 2010) in Chapter 4, and a connectionist model was used to represent the activations of concepts in working memory (Plummer, Wang, Tzeng, Pomplun, & Rayner, 2012) in Chapter 5.

My interests in reading and vision studies provided interdisciplinary research opportunities, which I pursued by applying methods and concepts from reading research to the viewing of real-world scenes. Regarding eye movements in natural scene viewing, I studied when and where we fixate, resulting in a model for gaze transition using LSA (Wang, Hwang, & Pomplun, 2010; Hwang, Wang, & Pomplun, 2011) as described in Chapters 6 and 7. Chapters 8 and 9 focus on studying how texts attract attention in natural scene viewing (Wang & Pomplun, 2012) compared to attraction by low-level visual features that are typically thought to induce saliency. For this purpose, I used experimental approaches and a computational model that includes an automatic text detector (Wang, Lu, Lim, & Pomplun, 2012). Finally, the conclusions regarding the general findings, implications of the results on linguistic and visual processing derived from the computational models, and practical applications are given in Chapter 10.

2. Using Singular Value Decomposition to Investigate Degraded Chinese Character Recognition: Evidence from Eye Movements During Reading

(under minor revision as Wang, Angele, Schotter, Yang, Simovici, Pomplun, & Rayner, *Journal of Research in Reading*)

It is known that not all letters are of equal importance to the word recognition process. Changes to initial letters of words are more disruptive to reading than are changes to medial or final letters (Rayner & Kaiser, 1975; Rayner, White, Johnson, & Liversedge, 2006). Furthermore, exterior letters are more important than word internal letters (Jordan, Patching, & Thomas, 2003; Rayner et al., 2006). Similar results were found in a previous study on stroke removal during Chinese reading, which indicates that removing initial strokes from Chinese characters makes them harder to read than removing final or internal ones (Yan, Bai, Zang, Bian, Cui, Qi, Rayner, & Liversedge, 2012). However, these studies of the Chinese writing system raise the question of whether there is something privileged about the first-written strokes or whether another aspect of the strokes at the beginning of the writing order is what causes them to be more important for character identification. One explanation is that the first written strokes can construct the visual configuration of a character quickly, and therefore facilitate successful character recognition without the presence of ending strokes. To test this, we turned to *Singular Value Decomposition* (*SVD*, Strang, 1993) to investigate whether the contribution of these strokes to the configuration of the character drives their importance for identification (see Figure 1).

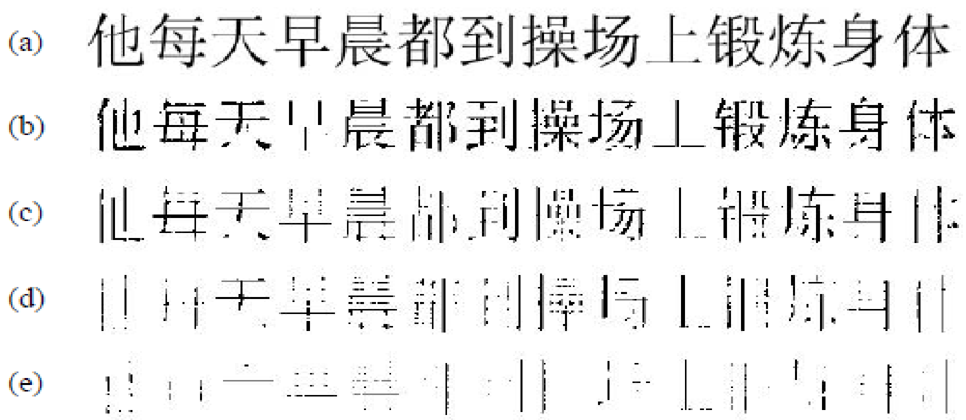


Figure 1: SVD reduction for Chinese characters. Sentence (a) is the original sentence without any reduction. Sentences (b) to (e) are the sentences after removing the least important 20%, 40%, 60%, and 80% information as determined by SVD.

In an eye-movement experiment, 48 sentences were presented in four experimental conditions: (1) all segments retained, (2) the least important 30% of segments removed, (3) the most important 30% of segments removed, and (4) 30% of segments randomly selected to be removed. The results were consistent with the Yan et al. (2012) study and indicated that when the least important segments—which did not seriously alter the configuration (contour) of a character—were deleted, subjects read as fast as when no segments were deleted. When the most important segments—which are located in the left side of a character and written first—were deleted, reading speed was greatly slowed.

These results suggest that SVD, which has no information about stroke writing order, can identify the most important strokes for Chinese character identification, i.e., mainly those that contribute to character configuration and contour (see Figure 2). Contour may be correlated with stroke writing order, which may lead to similarities between our data and the data pattern reported by Yan et al. (2012).

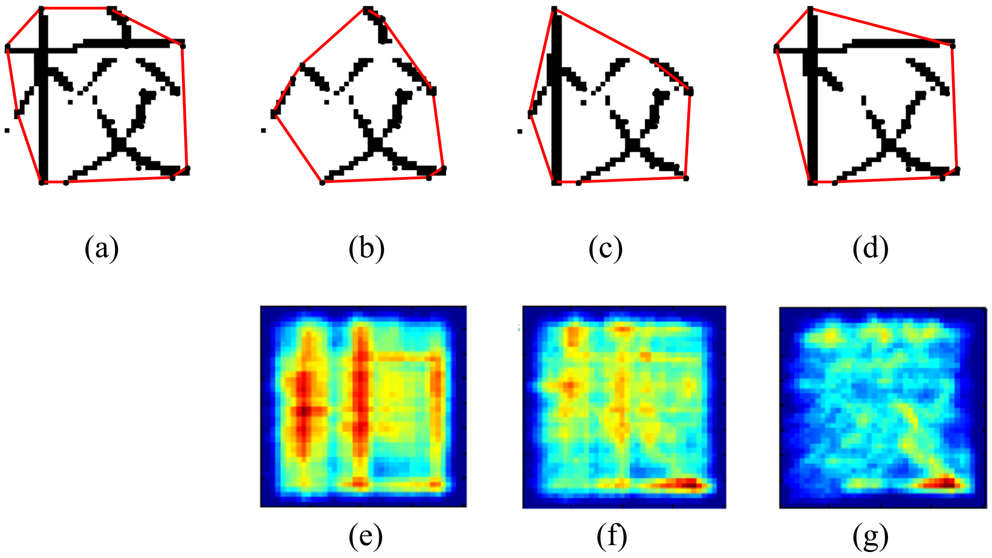


Figure 2: (a) to (d) are the contours of sample characters as defined by their convex hull. (a) All retained, (b) the most disruptive removal (most important segments), (c) moderately disruptive removal (randomly selected segments), and (d) the least disruptive removal (least important segments). (e) to (g): Distributions of removed segments (30%) of all characters, which were (e) the most, (f) moderately, and (g) the least disruptive.

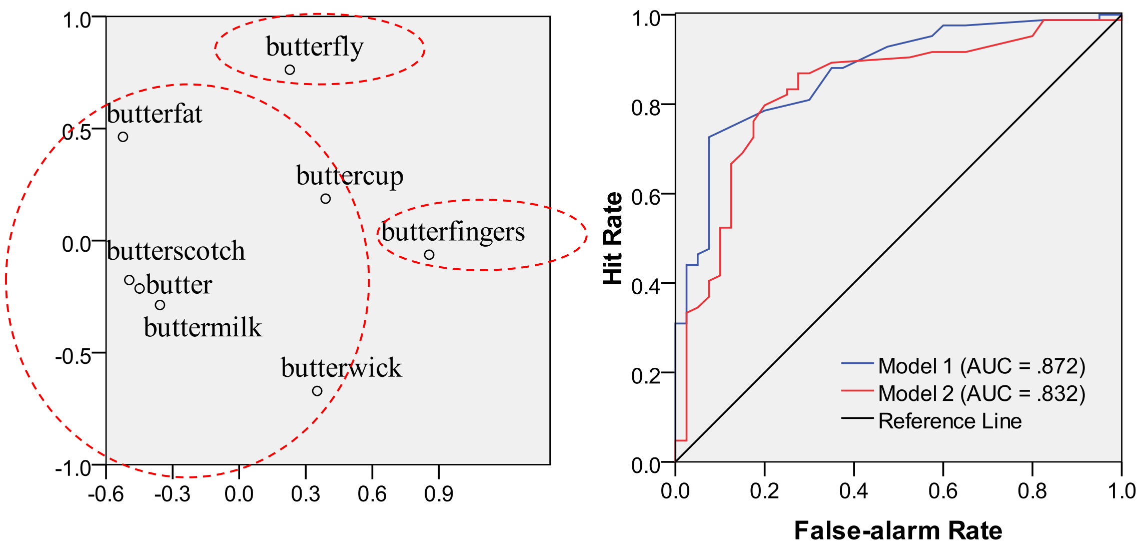
3. Predicting Raters' Transparency Judgments of English and Chinese Morphological Constituents using Latent Semantic Analysis Models

(parts presented at the Asia-Pacific Conference on Vision (APCV), 2010, and awarded a Student Traveling Grant (5 out of approximately 200 contributions); parts published as Wang, Tien, Hsu, & Pomplun, 2012, *CogSci proceedings*)

The morphological constituents of English compounds (e.g., butter and fly for butterfly) and two-character Chinese compounds may differ in meaning from that of the whole word that they form. However, judgments of semantic transparency are often subjective and vary strongly across raters, and a computational model may be a way to average across subjective differences.

Latent Semantic Analysis (LSA), the SVD-based method in linguistic studies, has been successful at simulating a wide range of psycholinguistic phenomena (see Jones & Mewhort, 2007, for a review). Therefore, LSA may be a solution to the problem of estimating semantic transparency of a compound word. The current chapter proposes two models based on this idea: Model 1 compares the semantic similarity between a compound word and each of its constituents, and Model 2 derives the dominant meaning of a constituent based on a clustering analysis of morphological family members (e.g., “butterfingers” or “buttermilk” for “butter”). The proposed models account for polysemy of constituents and successfully predicted participants’ transparency ratings, as shown in Figure 3.

Corroborating evidence from two different languages was presented by testing the stimuli used in prior compound word studies (Frisson, Niswander-Klement, & Pollatsek, 2008; Mok, 2009) as well as a rating experiment in the present study. Both Models 1 and 2 are predictive to the results of human transparent judgments, and the results indicate that Model 2 may in general be a better approach than Model 1 to predict transparency ratings for constituents of Chinese compounds. We propose that the models may explain the morphological processing when raters classify semantic transparency of English and Chinese compounds.



(a) (b)

Figure 3: (a) The MDS result for an example of semantic relationships for "butter" and its morphological family. The x and y axes represent dimensions 1 and 2, respectively, of the abstract, two-dimensional Euclidean output space of the MDS algorithm. (b) ROC analysis of Models 1 and 2 using the materials in Frisson et al. (2008).

4. Estimating the Effect of Word Predictability on Eye Movements in Chinese Reading using Latent Semantic Analysis and Transitional Probability

(published as Wang, Pomplun, Ko, Chen, & Rayner, 2010, *Quarterly Journal of Experimental Psychology*)

The preceding Chapter 3 has demonstrated that LSA-based models can explain the word identification process for both English and Chinese speakers. In the present chapter, we will be using LSA to address higher-level linguistic processing, namely the *predictability* effect between a target word and its prior context during sentence processing.

The predictability of target words (as typically determined by raters in a cloze task) has been found to strongly influence eye movements during reading (Rayner & Well, 1996; see also Ehrlich & Rayner, 1981). There are other computational methods that have been utilized to approximate predictability and its effect on eye movements, such as transitional probability (TP; see McDonald & Shillcock, 2003a; 2003b). McDonald and Shillcock found that TPs between words have a measurable influence on fixation durations and suggested that the processing system is able to draw upon statistical information in order to rapidly estimate the lexical probabilities of upcoming words. They also suggested that TP might reflect low-level predictability, which influences 'early' processing measures such as first fixation duration, instead of high-level predictability, which influences 'late' processing measures. The objective of the present study was to estimate word predictability, via the use of TP and LSA, and to further investigate predictability effects in Chinese reading when word complexity and frequency are taken into account.

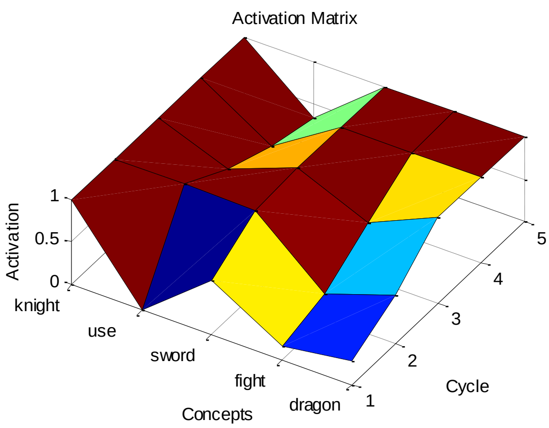
The results show that TP and LSA can be used as complementary tools for deriving word predictability ratings. Local information is retrieved by TP which considers only two consecutive words, while global information is utilized by LSA to bring out latent semantic relationships among words even if they have never co-occurred in the same document (Jones et al, 2007). In this sense, loosely speaking, TP reflects the word predictability effect on early stage lexical processing, while LSA reflects late stage processing during Chinese reading.

5. A Connectionist Model of Concept Activation during Reading using Latent Semantic Analysis and LandScape Model

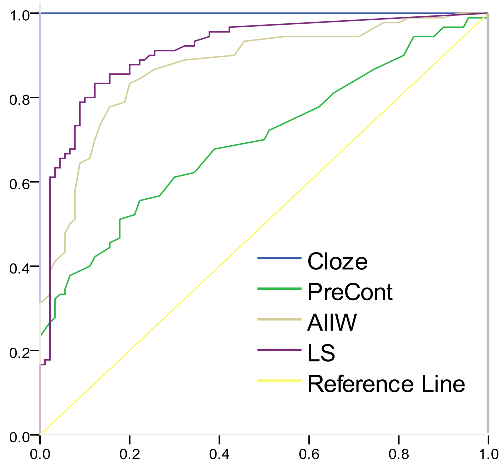
(parts published as Plummer, Wang, Tzeng, Pomplun, & Rayner, 2012, *CogSci proceedings*, and awarded a Student Travel Award , with 21 awards out of 798 contributions)

Using LSA and TP, we were able to estimate the results obtained by the cloze task and eye movements during reading. However, these computational methods did not explain how the semantic representation of each content word in a sentence is activated in working memory. In this chapter, we propose a connectionist model (Landscape model, see van den Broek, 2010) and LSA to determine the predictability of a word and its corresponding semantic representation associated in a neural network. LSA is used to establish connections between words and simulate the long-term semantic associations among concepts. This model may provide a means of investigating how language comprehension is affected by the activation of concepts in working memory (see an example in Figure 4a and the computation details in the dissertation).

We re-analyzed the materials from Gollan, Slattery, Goldenberg, Van Assche, Duyck, and Rayner (2011), in which predictable or unpredictable target word conditions were confirmed by a norming cloze task (Cloze). We estimated predictability of a target word by (1) the previous content word (PreCont), (2) all words in prior context (AllW), and (3) the estimates of the proposed connectionist model (LS) in this study. An ROC analysis demonstrates that the area under the curve (AUC) of Cloze, PreCont, AllW, and LS are 1, .70, .87, and .91, respectively, showing that the LS model obtains a higher AUC than AllW or PreCont (see Figure 4b). Furthermore, a correlation analysis demonstrates that the Pearson correlation coefficients between Cloze and PreCont, AllW, and LS are .39, .56, and .70, respectively. These results suggest that the LS model is superior over measures that utilize only the prior content word or LSA connections between content words exclusively. We suggest that modeling the process whereby linguistic inputs activate concepts in long-term memory and continuously influence working memory operations during sentence comprehension is an important endeavor in psycholinguistics.



(a)



(b)

Figure 4: (a) The "landscape" of the activation matrix for the Knight example. (b) ROC curves for Cloze, PreCont, AllW, and LS. The x and y axes represent false-alarm rate and hit rate, respectively.

6. Object Frequency and Predictability Effects on Eye Fixation Durations in Real-World Scene Viewing

(published as Wang, Hwang, & Pomplun, 2010, *Journal of Eye Movement Research*)

As indicated by the findings from Chapter 5, LSA can be used to estimate word predictability in reading. This successful application raises the question whether the perception of objects in everyday scene viewing can be studied in a similar way. During text reading, the durations of eye fixations decrease with greater frequency and predictability of the currently fixated word (Rayner, 1998; 2009). However, it has not been tested whether those results also apply to scene viewing. Therefore, we chose to base this study on an annotated image database, the freely available LabelMe image dataset (Russell, Torralba, Murphy & Freeman, 2008). The locations of objects are provided as coordinates of polygon corners and are labeled by English words or phrases. The given locations and labels of objects make the analysis of object-based eye-movement measures possible. Furthermore, the labels of English words provide an opportunity to apply linguistic-based methods, such as LSA, to scene viewing, and compare the results between vision and language. We computed object frequency and predictability from both linguistic and visual scene analysis, and LSA was applied to estimate predictability.

In a scene-viewing experiment, we found that, for small objects, linguistics-based frequency, but not scene-based frequency, had effects on first fixation duration, gaze duration, and total time. Both linguistic and scene-based predictability affected total time. Similar to reading, fixation duration decreased with higher frequency and predictability. For large objects, we found the direction of effects to be the inverse of those found in reading studies. These results suggest that the recognition of small objects in scene viewing shares some characteristics with the recognition of words in reading.

7. Semantic Guidance of Eye Movements in Real-world Scenes

(published as Hwang, Wang, & Pomplun, 2011, *Vision Research*)

In the previous chapter, we found that reading and scene viewing share some mechanisms, and the semantic effects estimated by LSA influence fixation durations. It is possible that semantic factors, such as meaning and semantic relations among objects, also affect where we look.

In order to estimate the effect of semantic similarities between objects based on visual scenes, the co-occurrence of objects in a large number of scene images and the importance of each object in the scene context - defined by its attributes such as size, location or luminance - would have to be carefully considered. It is important to notice, however, that semantic relations are formed at the conceptual rather than at the visual level and thus do not have to be derived from image databases. Consequently, any database that can generate a collection of contexts or knowledge might be used to represent the semantic similarity of objects.

For the present study, we chose LSA, the linguistics-based computational method, as a quantitative measure of semantic similarity between objects. Since annotated objects in LabelMe have descriptive text labels, their semantic similarity can be estimated by calculating cosine values for the labels of object pairs. We conducted two experiments to study two everyday visual activities - scene inspection and scene search. For each recorded eye fixation during scene inspection, we generated a semantic saliency map, which was entirely based on the semantic similarities between the currently fixated object and the other objects in the scene, as shown in Figure 5.

(a)

(b)

Figure 5: (a) Original scene image. (b) Semantic saliency map when the observer fixates on an object labeled as “FORK” (in the orange square); greater brightness indicates higher activation.

ROC analyses were carried out to evaluate the semantic saliency maps as predictors of subjects’ gaze transitions. With the controls of the random case, i.e., simulated random fixations in the same scene, and the dissociated case, i.e., the fixations from different scenes for the same subject, we found a preference that subjects’ eye movements were consistently guided toward objects that were semantically similar the current fixated one in scene inspection task, and progressively biased toward the objects semantically related to the search target in search task. These results indicate that there is substantial semantic guidance, defined by LSA, of eye movements in real-world scenes, which is not considered by any previous models of visual attention. Future research needs to address the importance of higher-level eye movement control in the context of semantic guidance, by using methods such as cross-linguistic investigation.

8. The Attraction of Visual Attention to Texts in Real-World Scenes

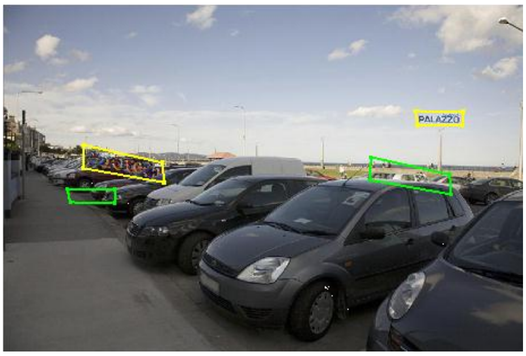
(published as Wang & Pomplun, 2012, *Journal of Vision*)

The results of Chapter 7 indicated that semantic factors affect where we look, which raises questions regarding how people process texts in real-word scenes. For instance, how do people locate and read signs or billboards that are embedded in a complex environment? Do semantic factors affect how fast we access texts, e.g., words vs. scrambled words, or English vs. Chinese texts for English vs. Chinese speakers?

Texts in real-world scenes were found to attract more attention than regions with similar size and position in a free viewing task (Cerf, Frady, & Koch, 2009), but it is still an open question what factors would control such an attentional bias toward texts. It is possible that low-level visual saliency attracts attention (e.g., Itti, Koch & Niebur, 1998; Bruce & Tsotsos, 2006; Itti & Koch, 2001; Parkhurst, Law, & Niebur, 2002). It is also possible that the typical *locations* of texts in the scene context are more predictable to contain important information, which would be in line with the contextual guidance model (Torralba, Oliva, Castelhano, & Henderson, 2006), scene syntax (Võ & Henderson, 2009), and dependency among objects (Oliva & Torralba, 2007; Mack & Eckstein, 2011). Finally, the observer's *familiarity* with texts, i.e., either low-level visual features of a specific writing system or text semantics, might influence the attractiveness of texts. The goal of the present study was to investigate the contributions of low-level visual saliency, expected locations, specific visual features, and familiarity of texts to their ability to attract attention in real-world scene viewing. As shown in Figure 6a, in order to test if texts (in yellow polygon) are more attractive than other scene objects (in green polygon), in Experiment 1 an eye-tracking database of scene viewing by Judd, Ehinger, Durand, and Torralba (2009) was first reanalyzed.

(a)

(b)



(c)

(d)

Figure 6: A sample stimulus. The paired control regions are shown in green polygons. (a) Texts (yellow polygons) in Experiment 1. (b) Erased texts (yellow polygons) in Experiment 2. (c) Unconstrained texts (yellow polygons) placed in front of homogeneous (right) and inhomogeneous backgrounds (left) in Experiment 3. (d) Words (yoyo) and drawings (sled) on homogeneous background. There are four versions of stimuli paired either a regular word or scrambled word with a drawing.

In Experiments 2 to 5 (see Figures 6b, 6c, 6d, and 7), new eye-movement data were collected and analyzed to study the factors of expected location, text features, semantics, and familiarity underlying the attraction of attention by texts.

(a) (b)

Figure 7: Example of upside-down and Chinese texts used in Experiment 5. (a) Version C1, in which half of the original texts were rotated and the other half was replaced with Chinese texts. (b) Version C2, in which the upside-down texts in C1 were replaced with Chinese texts, and the Chinese texts in C1 were replaced with upside-down texts.

9. Visual Attention is Attracted by Text Features Even in Scenes without Text

(published as Wang, Lu, Lim, & Pomplun, 2012, *CogSci proceedings*)

As we have seen in the preceding chapter, viewers' attention is disproportionately attracted by texts, especially by the ones they are familiar with. A possible reason is that viewers have *developed* a "text detector" in their visual system to bias their attention toward some specific text features. One way to verify this hypothesis is to add a text detector module to a visual attention model and test if the inclusion increases the model's ability to predict eye fixation positions. In a previous study, adding a module of manually-defined regions of texts was shown to improve the prediction of eye fixations in text-present images (Cerf et al., 2009). However, it is still unclear if viewers' attention is biased toward any non-text objects which share some features of texts, particularly in text-absent images. Therefore, an *automatic* text detector based on the recognition of specific text features is required to address this question.

We found that adding a text detector to an attention model improved its prediction of viewers' visual attention, even in text-absent images. Our results suggest that non-text objects whose features resemble those of texts (such as high spatial frequency edges) catch a disproportionate share of attention. Based on the current data, it seems that the viewers' "biological text detectors" are somewhat similar to the artificial system and influence the viewers' distribution of attention when viewing real-world images. From a time-course analysis, it appears that the biological text detector influences the allocation of attention particularly strongly during later stages of image inspection when viewers are increasingly likely to attend to detailed local structures (see Unema et al., 2005) for semantic interpretation of perceived text.

10. Conclusions

My research has an interdisciplinary focus in many perspectives. My research topics include vision science and language processing, from low-level visual processing, attention, to high-level semantics in memory, as well as cross-linguistic investigations. My research paradigms are experimental (eye-tracking) studies and computational modeling. I used novel interdisciplinary approaches by applying computational techniques such as SVD, LSA, connectionist modeling, computational linguistics, and computer vision, to the problems of reading and scene-viewing.

In a highly interdisciplinary field such as the modeling of human eye movements during cognitive tasks, collaboration with other researchers from various departments is important, which I believe led to stronger, more insightful results and a better learning experience for all researchers involved. Much of the work reported in this thesis was carried out collaboratively and resulted in journal or conference proceedings publications.

The findings should eventually lead to practical applications, e.g., contribute to the development of more effective automatic text detectors, or making a great difference to visually challenged people's lives by assisting them in reading and scene viewing. After completing my thesis, I have been awarded a postdoctoral fellowship in the Robotics, Vision, and Sensor Networks Group (RVSN), Computer Science and Artificial Intelligence Laboratory (CSAIL), at MIT to work with Dr. Seth Teller on a "text spotting" project on wearable devices for visually impaired people. The project required interdisciplinary expertise such as computer vision (text detection and tracking), degraded character recognition, and computational linguistics (word recognition), which were well-developed throughout my Ph.D. training.

Taken together, the results of my doctoral thesis improve our understanding of low-level and higher-level cognitive processing as well as cultural differences during reading and real-world scene viewing, which provide insightful opportunities to broaden the relevant research areas. I therefore believe that my thesis is well aligned with the principles of the Robert J. Glushko Dissertation Prize in Cognitive Science.

References

Bruce, N. D. B., & Tsotsos, J. K. (2006). Saliency based on information maximization. *Advances in Neural Information Processing Systems*, 18, 155–162.

Cerf, M., Frady, E. P., & Koch, C. (2009). Faces and text attract gaze independent of the task: Experimental data and computer model. *Journal of Vision*, 9(12):10, 1–15.

Ehrlich, S.F., & Rayner, K. (1981). Contextual effects on word perception and eye movements during reading. *Journal of Verbal Learning and Verbal Behavior, 20*, 641-655.

Frisson, S., Niswander-Klement, E., & Pollatsek, A. (2008). The role of semantic transparency in the processing of English compound words. *British Journal of Psychology*, 99, 87-107.

Gollan, T. H., Slattery, T. J., Goldenberg, D., Van Assche, E., Duyck, W., & Rayner, K. (2011). Frequency Drives Lexical Access in Reading but Not in Speaking: The Frequency-Lag Hypothesis. *Journal of Experimental Psychology: General*,140, 2, 186–209.

Hwang, A. D., Wang, H. C., & Pomplun, M. (2011). Semantic guidance of eye movements in real-world scenes. *Vision Research*, 51, 1192-1205.

Itti, L, Koch, C., & Niebur, E. (1998). A Model of Saliency-Based Visual Attention for Rapid Scene Analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 20 (11): 1254-1259.

Itti, L., & Koch, C. (2001). A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Research*, 40(10-12), 1489-1506.

Jones, M. N. & Mewhort, D. J. K. (2007). Representing word meaning and order information in a composite holographic lexicon. *Psychological Review*, 114, 1-37.

Jordan, T. R., Thomas, S. M., Patching, G. R., & Scott-Brown, K. C. (2003). Assessing the importance of letter pairs in initial, exterior, and interior positions in reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 29(5), 883-393.

Judd, T., Ehinger, K., Durand, F., & Torralba, A. (2009). Learning to predict where humans look, *IEEE International Conference on Computer Vision (ICCV)*, Kyoto, Japan, 2106 - 2113.

Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review, 104,* 211–240.

Mack, S. C., & Eckstein, M. P. (2011). Object co-occurrence serves as a contextual cue to guide and facilitate visual search in a natural viewing environment. *Journal of Vision*, 11(9):9, 1-16.

McDonald, S. A., & Shillcock, R. C. (2003a). Eye movements reveal the on-line computation of lexical probabilities during reading. *Psychological Science, 14*, 648–652.

McDonald, S. A., & Shillcock, R. C. (2003b). Low-level predictive inference in reading: The influence of transitional probabilities on eye movements. *Vision Research, 43*, 1735–1751.

Mok, L. W. (2009). Word-superiority effect as a function of semantic transparency of Chinese bimorphemic compound words. *Language and Cognitive Processing*, 24 (7/8), 1039-1081.

Oliva, A., & Torralba, A. (2007). The role of context in object recognition. *Trends in Cognitive Science*, 11, 520–527.

Parkhurst, D. J., Law, K., & Niebur, E. (2002). Modeling the role of salience in the allocation of overt visual selective attention. *Vision Research*, 42, 107–123.

Plummer, P., Wang, H. C., Tzeng, Y. T., Pomplun, M., & Rayner K. (2012). A Connectionist Model of Concept Activation during Reading using Latent Semantic Analysis and LandScape Model. In *Proceedings of Annual Meeting of the Cognitive Science Society (CogSci 2012)*, Sapporo, Japan.

Rayner, K. (1998). Eye movement in reading and information processing: 20 years of research. *Psychological Bulletin, 24,* 372-422.

Rayner, K. (2009). The 35th Sir Frederick Bartlett Lecture: Eye movements and attention in reading, scene perception, and visual search. *Quarterly Journal of Experimental Psychology, 62*, 1457-1506.

Rayner, K., & Kaiser, J. S. (1975). Reading mutilated text. *Journal of Educational Psychology*, *67*, 301-306.

Rayner, K., & Well, A. D. (1996). Effects of contextual constraint on eye movements in reading: A further examination. *Psychonomic Bulletin & Review, 3*, 504–509.

Rayner, K., White, S., Johnson, R. & Liversedge, S. (2006). Raeding Wrods With Jubmled Lettres: There Is a Cost. *Psychological Science, 17,* 192-193.

Russell, B. C., Torralba, A., Murphy, K. P. & Freeman, W. T. (2008), LabelMe: a database and web-based tool for image annotation, *International journal of computer vision, volume 77, issue1-3*, 157-173.

Strang, G. (1993). Introduction to Linear Algebra, 2nd Edition, Wellesley-Cambridge Press.

Torralba, A., Oliva, A., Castelhano, M., & Henderson, J.M. (2006). Contextual guidance of eye movements and attention in real-world scenes: the role of global features in object search. *Psychological Review,* 113, 766-786.

Unema, P. J. A., Pannasch, S., Joos, M., & Velichkovsky, B.M. (2005). Time course of information processing during scene perception. *Visual Cognition*, 12(3), 473-494.

van den Broek, P. (2010). Using texts in science education: cognitive processes and knowledge representation. *Science*, 328, 453.

Võ, M. L.-H., & Henderson, J. M. (2009). Does gravity matter? Effects of semantic and syntactic inconsistencies on the allocation of attention during scene perception. *Journal of Vision*, 9(3):24, 1-15.

Wang, H.-C., & Pomplun, M. (2012). The attraction of visual attention to texts in real-world scenes. Journal of Vision, 12(6):26, 1–17.

Wang H. C., Hwang, A. D. & Pomplun, M. (2010). Object frequency and predictability effects on eye fixation durations in real-world scene viewing. *Journal of Eye Movement Research*, 3(3):3, 1-10.

Wang, H. C., Pomplun, M., Ko, H. W., Chen M. L., & Rayner, K. (2010). Estimating the effect of word predictability on eye movements in Chinese reading using latent semantic analysis and transitional probability, *Quarterly Journal of Experimental Psychology*, 63, 1374-1386.

Wang, H. C., Tien, Y. M., Hsu, L. C., & Pomplun, M. (2012). Estimating Semantic Transparency of Constituents of English Compounds and Two-Character Chinese Words using Latent Semantic Analysis. In *Proceedings of Annual Meeting of the Cognitive Science Society (CogSci 2012),* Sapporo, Japan.

Wang, H. C., Lu, S. J., Lim, J. H., & Pomplun, M. (2012). Visual Attention is Attracted by Text Features Even in Scenes without Text. In *Proceedings of Annual Meeting of the Cognitive Science Society (CogSci 2012)*, Sapporo, Japan.

Wang, H. C., Schotter, E., Angele, B., Yang, J. M., Simovici, D., Pomplun, M., & Rayner, K. Using Singular Value Decomposition to Investigate Degraded Chinese Character Recognition: Evidence from Eye Movements During Reading. Manuscript in minor revision in *Journal of Research in Reading*, *special issue on Chinese reading*.

Yan, G., Bai, X., Zang, C., Bian, Q., Cui, L., Qi, L., Rayner, K. & Liversedge, S. (2012). Using stroke removal to investigate Chinese character identification during reading: evidence from eye movements. *Reading and Writing,* 25, 951-979.