Australian National University

Predicting reading comprehension and dynamic text presentation in eLearning using eye gaze

Tom Gedeon

Research School of Computer Science, ANU

tom@cs.anu.edu.au

iHcc Information and Human Centred Computing

Australian National University

Outline

- Motivation
- Background
- Method
- Results
- Implications
- Conclusion

Australian National University

Can we predict reading comprehension?

- eText is ubiquitous
- eLearning is becoming ubiquitous
- Predicting comprehension removes need to consistently explicitly assess students
- Adaptive content generation based on predictions
- Prediction from eye movements is difficult
  - How do we improve predictions?
  - Use a flexible error function for training?

Australian National University

How Do People Read?

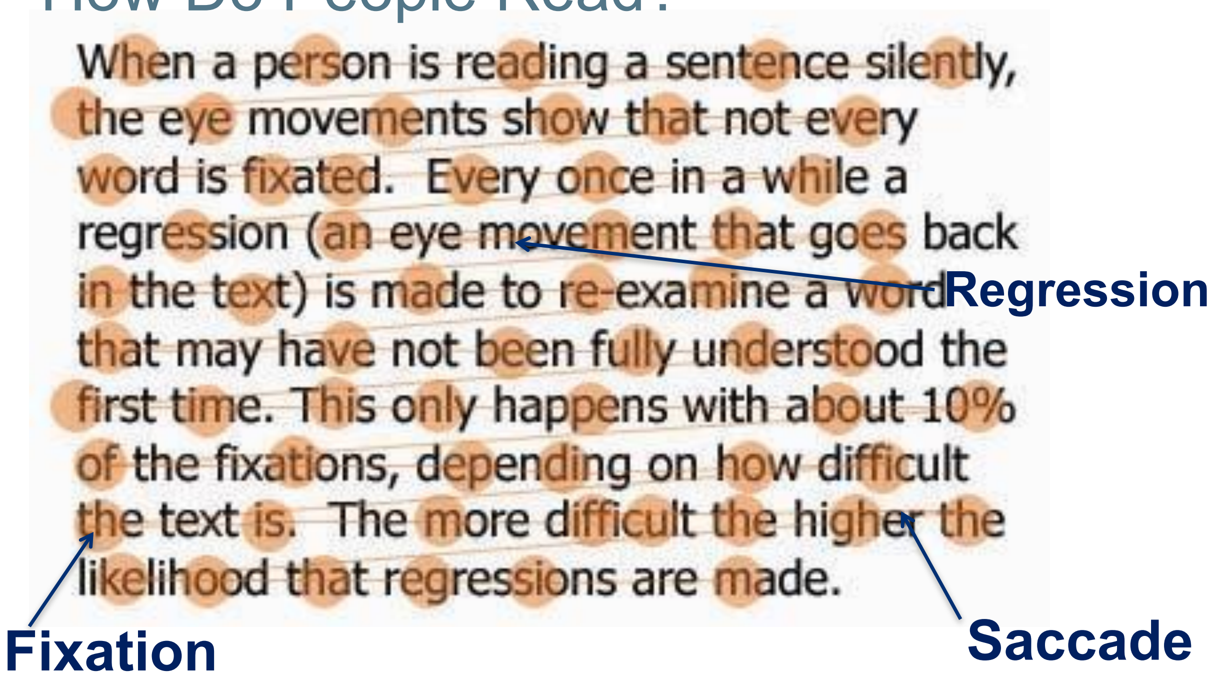


Image taken from: http://www.scholarpedia.org/article/File:Reading.jpg

How do we capture eye movements?

- Eye tracking!
  - Capture where the eye is looking
- We use video-based tracking
  - Non-intrusive,
  - Infrared light projected at the reader, reflects off the eye and sensed by special cameras
  - Typically corneal reflection and the centre of the pupil is used to track the eye and calculate eye rotation

Predicting Reading Comprehension

- Correlation between fixation duration and comprehension (Underwood, et al. 1990)
- Correlation between regressions and text difficulty (Rayner, et al., 2006) .
- Random forests and SVMs were used to predict readers understanding and language skill from multiple eye movement measures (Martínez-Gómez and Aizawa, 2014)
  - Good prediction of high and low performers but not of quantification of readers understanding

Eye Tracking in Adaptive eLearning

- Reading assistance
  - iDict (Hyrskykari et al., 2000)
  - The Reading Assistant (Sibert et al., 2000)
- Eye gaze can be used to differentiate the types of content being read, using SVMs and ANNs (Vo et al., 2010)
- ANNs have been used to predict item difficult in multiple-choice questions (Perkins et al. 1995)

Past Research

- Different sequences of text and questions shown to participants
  - All text were the same level of difficulty (easy to read)
- Artificial neural networks used to predict reading comprehension scores
  - Custom error function: Fuzzy Output Error (FOE)
  - FOE describes the error in a fuzzy way and then sums the fuzzy errors together to get the total error.

Discussion

- Difficult prediction problem
  – Best classification for 3 hidden layer networks
- Formats with questions and text on the same page generated the best results (~90% correct classification)
- When only text is shown to participants only chance results are achieved
  – Complicated relationship between eye movements and understanding
  – Need for improvement in classification for this case

Limitations of this study

- Is the text difficulty aligned with perceptions of difficulty?
  – Could result in poor predictions
- The text had the same level of difficulty
  – how to does changing the difficulty change predictions and perceptions of difficulty?
- Are there differences in perceptions between L1 and L2 readers?
  – Create separate classifiers for L1 and L2 readers

Can we improve predictions of reading comprehension?

- One method is by using better prediction techniques
- Another, is by exploiting the fact that text difficulty affects eye movements
  - Difficult text = more fixations, more regressions (Rayner et al., 2006)
  - Does making text more difficult to read make it easier to detect comprehension from eye movements?

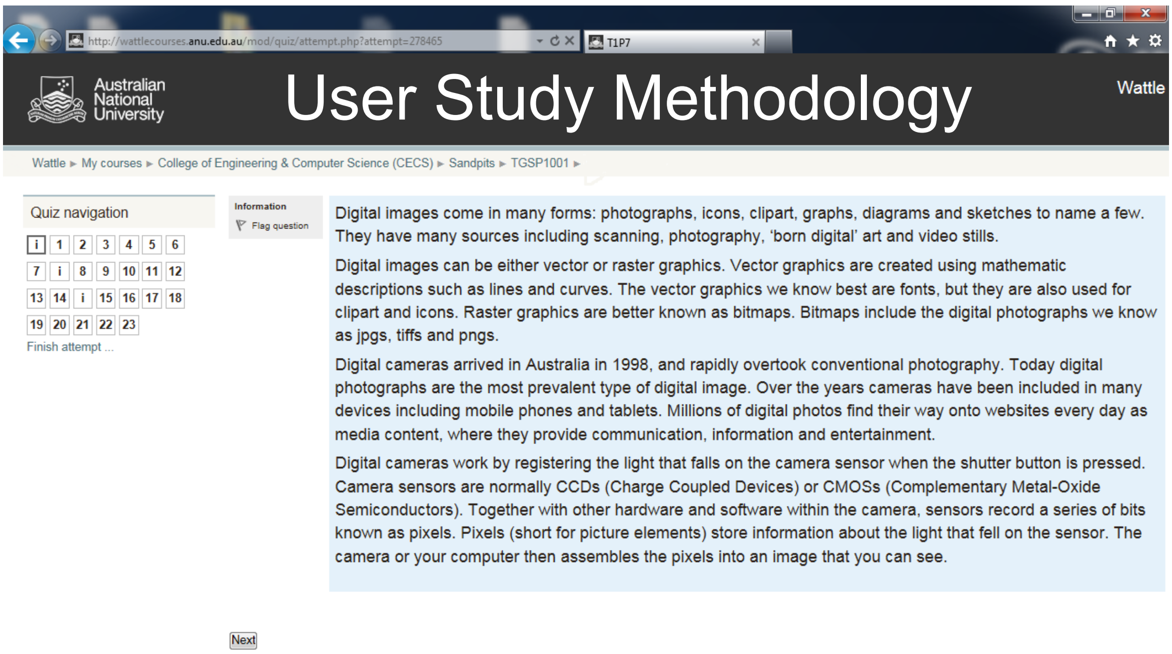


Figure 1. Example of text shown in Wattle eLearning Environment

9 levels of text difficulty

	Readability		
Concept	**Easy**	**Medium**	**Difficult**
Easy	A	B	C
Medium	D	E	F
Difficult	G	H	J

Figure 2. Description of the text property breakdown.

- To test for generalisability there are 3 sets of 9 texts
- Each set covered a different topic on the main topic of Digital images:
  - Working with Digital Images,
  - Copyright and Intellectual Property
  - Photo Credibility

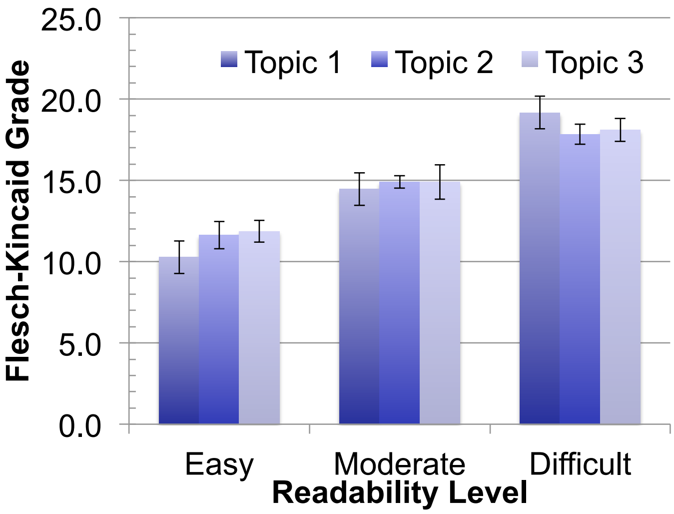


Figure 3. The Flesch-Kincaid readability grade level broken down for each level of readability and topic.

Paths

- Each participant was given a sequence of texts to read.
    1. A B D
    2. A E J
    3. A B G
    4. A D C
    5. A E H
    6. A E F
    7. A D B

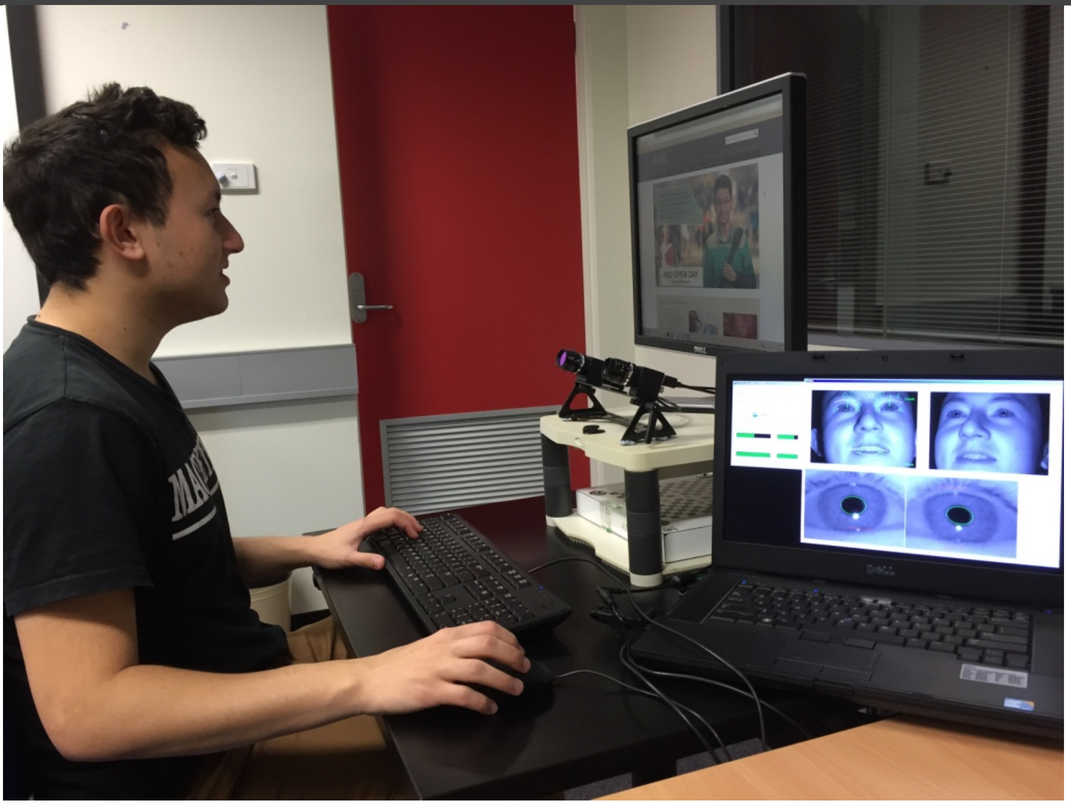
Questionnaire

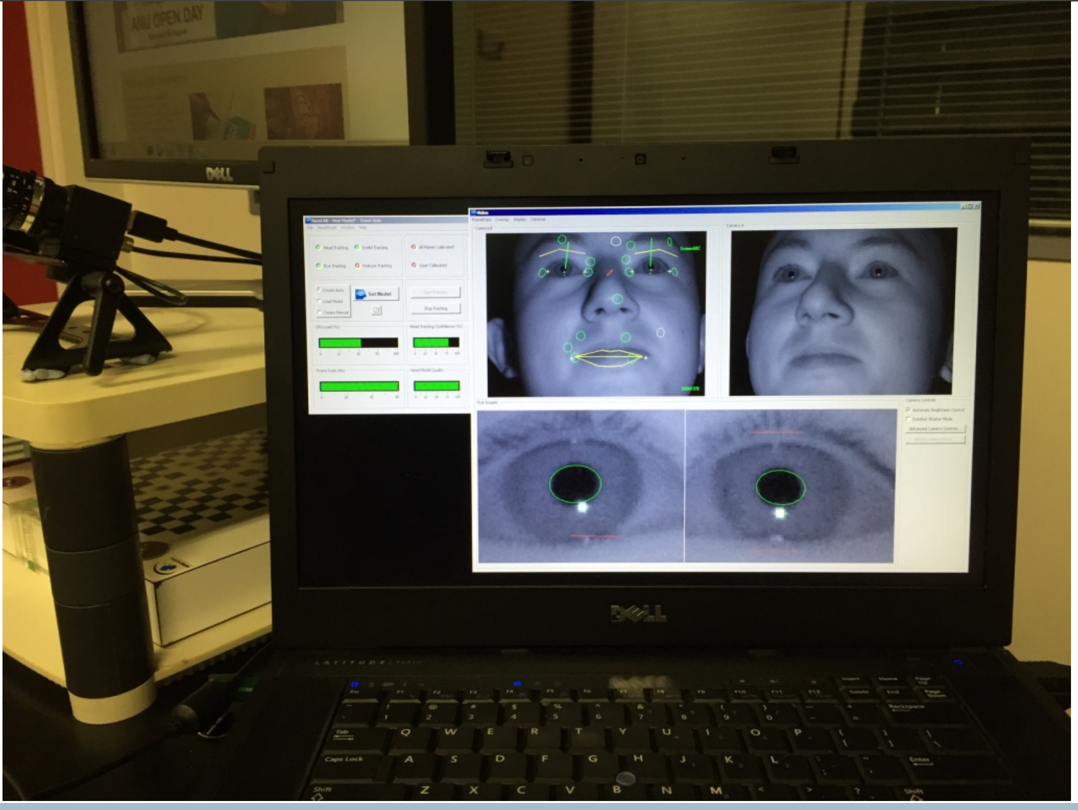
- How well do you think you understood the text? *(Very well / Well / Somewhat / Not at all)*
- How confident were you answering the questions? *(Very confident / Confident / Not Confident)*
- How difficult did you find the text to read? *(Easy / Moderate / Hard)*
- How complex was the concept being explained in the text? *(Basic / Intermediate / Advanced)*

Experiment setup

- Displayed to participants through online learning environment used at ANU
  - Wattle (Moodle variant)
- displayed on a 1280x1024 pixel Dell monitor.
- Eye gaze recorded at 60Hz using Seeing Machines FaceLAB 5 infrared cameras mounted at the base of the monitor
  - This eye tracker has a gaze direction accuracy of 0.5-1° rotational error and measures pupil diameter as well as blink events.
  - 9-point calibration prior to data collection for each participant

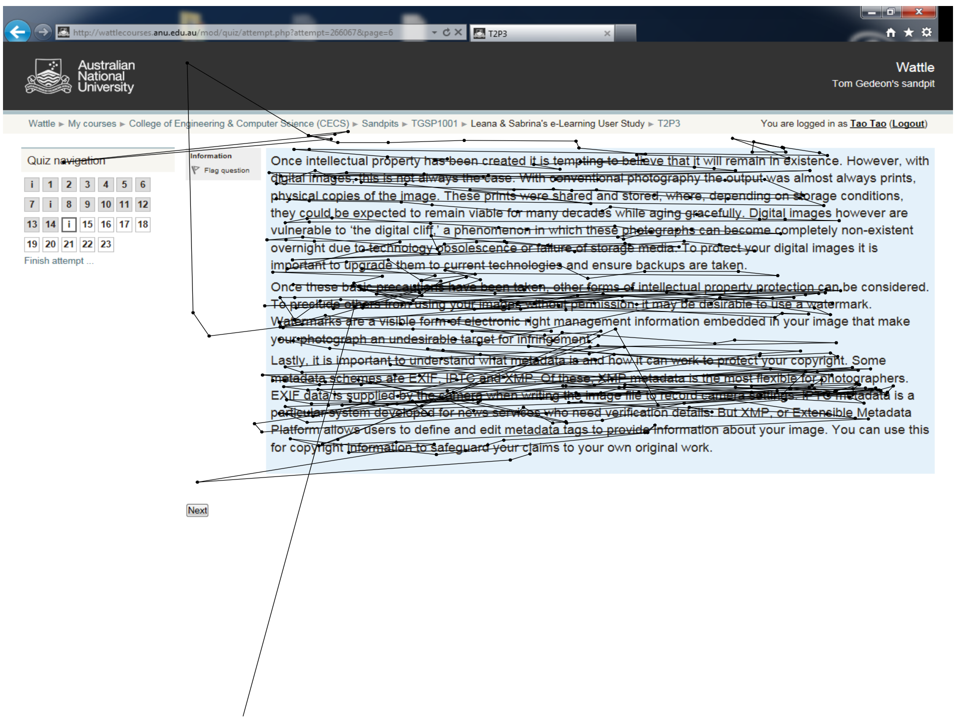
Australian
National
University




ANU OPEN DAY

Participants

- The eye gaze of 70 participants (47 male, 23 female) was recorded. Participants had an average age of 25 years (9 years standard deviation, range of 18 to 60 years).
  - 46 stated that English was the first language
  - 24 stated a language other than English



Australian National University

Data Pre-prosessing

- Eye gaze (x,y-coords) converted to fixations
- Fixations converted to eye movement measures:
  - *Normalised Num. fixations, Max fixation dur (s), Ave. fixation dur (s), Normalised total fixation dur (s), regression ratio, Ave. forward saccade, Reading analysis, Regional Analysis, Distractions counted*
- Calculated for each page

Australian National University

Research Question:

- Can we improve predictions of reading comprehension through manipulations of text readability and conceptual difficulty?

Hypotheses:

1. Better prediction will be obtained for L1 readers;
2. Better results will be obtained when the text is most difficult; and,
3. Predictions will be consistent between different versions of texts with the same complexity.

Predictions: L1 readers only

Text ID	**Text Properties**		**Topic 1**	**Topic 2**	**Topic 3**
	Readability	Concept			
A	Easy	Basic	0.34	0.21	0.35
B	Moderate	Basic	0.34	0.28	0.29
C	Difficult	Basic	0.50	0.24	**0**
D	Easy	Intermediate	0.32	0.40	0.13
E	Moderate	Intermediate	0.56	0.41	0.17
F	Difficult	Intermediate	0.62	**0**	0.62
G	Easy	Advanced	**0.06**	0.42	**0**
H	Moderate	Advanced	**0.07**	1.22	0.52
J	Difficult	Advanced	**0.13**	0.41	**0.12**
	Average		0.33	0.40	0.25

Predictions: L2 readers only

Text ID	Text Properties		Topic 1	Topic 2	Topic 3
	Readability	Concept			
A	Easy	Basic	0.41	0.39	0.30
B	Moderate	Basic	0.38	0.32	0.27
C	Difficult	Basic	1.04	1.45	0.43
D	Easy	Intermediate	0.68	0.20	0.40
E	Moderate	Intermediate	0.39	0.42	0.23
F	Difficult	Intermediate	0.51	0.25	0.50
G	Easy	Advanced	0.51	1.04	0.49
H	Moderate	Advanced	0.29	1.03	-
J	Difficult	Advanced	**0.11**	**0.09**	1.00
	Average		0.48	0.58	0.45

Summary

- Hypothesis 1: validated
  - L1 predictions are better than L2 predictions
- Hypotheses 2 and 3: no generalisation but there is a trend in better predictions for the more difficult texts as we hypothesised
- Why is there no generalisation?
  - Need to check now that the eye movements are as we would expect them to be…

Eye Movements – Norm. num. fixations

Text ID	Topic 1		Topic 2		Topic 3	
	L1	L2	L1	L2	L1	L2
A	0.68	0.84	0.62	0.79	0.68	0.79
B	0.66	0.93	0.78	0.72	0.72	0.75
C	0.9	0.75	0.67	0.74	0.75	1.04
D	0.66	0.74	0.76	0.91	0.7	0.76
E	0.78	1.01	0.6	0.89	0.79	1.04
F	1.23	1.3	0.67	0.56	0.88	0.69
G	0.71	1.13	0.7	0.9	0.43	0.99
H	0.82	0.77	0.79	1.0	0.69	0.6
J	1.06	1.16	0.73	0.97	0.89	1.05

Australian National University

Effect of Text on NNF

- For L1 readers:
  - for both Topics 1 and 3 there is an increase in NNF as readability increases (expected from past research)
  - Increase also seen as conceptual level increases!
  - more pronounced for Topic 1
  - No pattern in NNF for Topic 2
- For L2 readers:
  - higher NNF values (expected from past research)
  - Pattern of increased NNF due to text difficulty not clear for L2 readers

Implications

- A step forward in prediction of reading comprehension
- Intended use in adaptive eLearning:
  - omission of some assessment questions
  - differentiate between actual understanding and accidental choice of the correct answer
- Using standard readability formula, (e.g. the Flesch-Kincaid Grade Level) to assess the readability of text is not sufficient

Interim summary

- Predictions for the L1 dataset better for the L2 dataset
- Text properties have an effect on the predictions rates, however:
  - not consistent between different topics
  - investigation of eye movements reveals this is due to the eye movements not matching expected text difficulty in all topics
  - for the topics where the eye movements conformed to text difficulty, there is a trend of lower MSEs when the concept level is advanced.

Where to now?

- Since the eye movements do not reflect the intended difficulty for all of the text we have to ask:
  - Are the texts perceived as being that difficult?
- Investigation of the qualitative data collected

Remember: Limitations of past study

- Is the text difficulty aligned with perceptions of difficulty?
  - Could result in poor predictions
- The text had the same level of difficulty
  - how to does changing the difficulty change predictions and perceptions of difficulty?
- Are there differences in perceptions between L1 and L2 readers?
  - Create separate classifiers for L1 and L2 readers

We Investigate these limitations now

Perceptions of Text Complexity

- Participants were asked to rate the difficulty of the text in a questionnaire after reading each piece of text
- Given that the eye movements do not align exactly with how we would expect them to this raises the question of whether the text is actually as difficult as it is meant to be?

Hypotheses

1. Perceptions of text difficulty between the L1 and L2 readers will be different
2. Changes in text difficulty will be reflected by changes in perceived difficulty, however,
   a. the readability and conceptual difficulty will interplay to cause deviations of perceptions from the expected difficulty,
3. Eye movements can be used to provide a more accurate rating system for text difficulty

Perception results

	Readability		
Concept	**Easy**	**Medium**	**Difficult**
Easy	L2: under estimate A	Both under estimate B	Both under estimate C
Medium	D	Both over estimate E	L2: under estimate F
Difficult	Both under estimate G	H	L2: over estimate J

Figure 4. Visualisation of perceptions of text complexity

Interim Summary

- L1 and L2 readers have different perceptions of difficulty
- Readability and Concept level interact
  - how distinguishable is readability from conceptual level?
- Implication:
  - readability and conceptual level can be manipulated to effect perceived overall difficulty

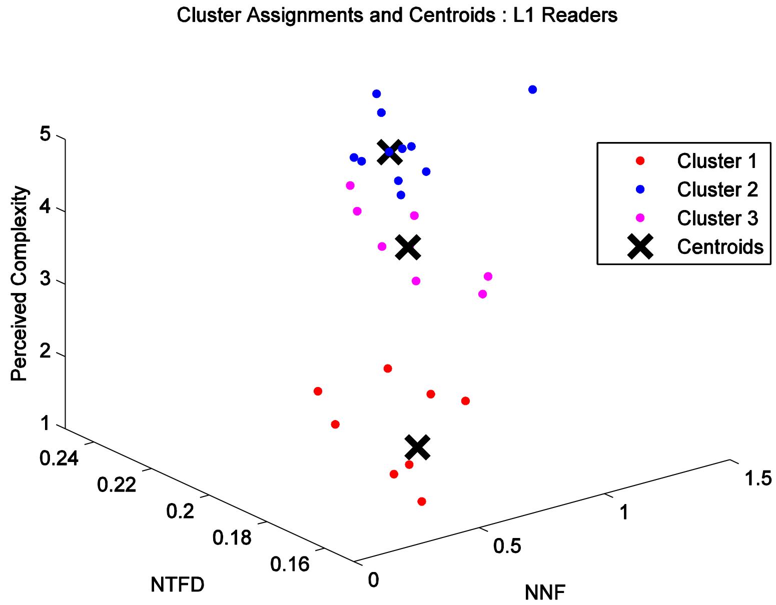
Australian National University

Now to the last hypothesis

- eye movements can be used to provide a more accurate rating system for text difficulty
- Using cluster analysis of the eye movement measures can we rank the texts in difficulty

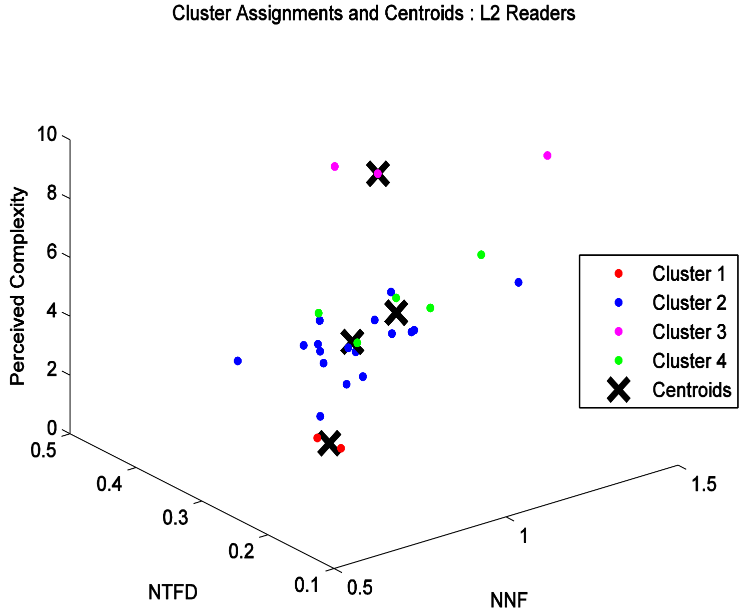
Australian National University

Ranking Perceived Text Complexity using eye movements: L1 Readers



Australian National University

Ranking Perceived Text Complexity using eye movements: L2 Readers



Australian National University

Implications

- The cluster analysis reveals two different difficulty ratings
  - One for L1 reader, the other for L2 readers
- Gives median perceived difficulty of text
  - This can be used to gauge the perceived difficulty
  - Teachers can use this information to produce better learning materials in eLearning
- Use in adaptive eLearning:
  - Find the difficulty of the text as perceived by each student and change the text given to them based on this rating – Better targeting of skill level

Overall Discussion

- The text properties have a bearing on prediction results
  - Making a text more complex will increase prediction results
  - However, text actually has to be complex; cannot rely on traditional readability statistics
- Readability and conceptual difficulty interact to distort overall perceptions
- Eye movements can be used to get average perceived text complexity
  - There are clear difference between L1 and L2 readers

Further Work

- More work to be done on getting better predictions
- Consider pupil dilation, GSR, ECG, EEG, etc.
- Look at different types of texts, do these affect the eye movements the same?
  - e.g. highly technically
- Validating if adaptive changes in eLearning provide benefit
  - short and longitudinal study
- Mobile environments e.g. tablets, smartphones

References (if you can't download a paper, email: tom@cs.anu.edu.au)

Kim, J., Thomas, P., Sankaranarayana, R., Gedeon, T., & Yoon, H. J. (2015). Eye-tracking analysis of user behavior and performance in web search on large and small screens. *Journal of the Association for Information Science and Technology*, 66(3), 526-544.

Copeland, L., Gedeon, T., & Mendis, S. (2014). Predicting reading comprehension scores from eye movements using artificial neural networks and fuzzy output error. *Artificial Intelligence Research*, *3*(3), p35.

Copeland, L., & Gedeon, T. (2014, December). Effect of presentation on reading behaviour. In *Proceedings of the 26th Australian Computer-Human Interaction Conf. on Designing Futures: the Future of Design* (pp. 230-239). ACM.

Copeland, L., Gedeon, T., & Mendis, S. (2014). Fuzzy Output Error as the Performance Function for Training Artificial Neural Networks to Predict Reading Comprehension from Eye Gaze. *In Neural Information Processing* (pp. 586-593). Springer International Publishing.

Copeland, L., Gedeon, T., & Mendis, S. (2014). Predicting reading comprehension scores from eye movements using artificial neural networks and fuzzy output error. *Artificial Intelligence Research*, 3(3), p35.

Dhall, A., Goecke, R., Joshi, J., Sikka, K., & Gedeon, T. (2014, November). Emotion recognition in the wild challenge 2014: Baseline, data and protocol. In *Proceedings of the 16th International Conference on Multimodal Interaction* (pp. 461-466). ACM.

Gregor, S., Lin, A., Gedeon, T., Riaz, A., Zhu, D. (2014) Neuroscience and a nomological network for the understanding and assessment of emotions in information systems research, *Journal of Management Information Systems (JMIS)*, 30(4), 13-48.

Sharma, N., and Gedeon, T.D. (2014) Modeling a stress signal, *Applied Soft Computing*, 14, 53-61.

Copeland, L., & Gedeon, T. (2013, December). Measuring reading comprehension using eye movements. In *Cognitive Infocommunications (CogInfoCom), 2013 IEEE 4th International Conference on* (pp. 791-796). IEEE.

Copeland, L., & Gedeon, T. (2013, November). The effect of subject familiarity on comprehension and eye movements during reading. In *Proceedings of the 25th Australian Computer-Human Interaction Conference: Augmentation, Application, Innovation, Collaboration* (pp. 285-288). ACM.

Australian National University

References continued

Sharma, N., and Gedeon, T. (2013) Computational Models of Stress in Reading Using Physiological and Physical Sensor Data, In *Advances in Knowledge Discovery and Data* Mining, pp. 111-122, Springer Berlin Heidelberg.

Sharma, N. and Gedeon, T. (2013) Modeling stress recognition in typical virtual environments, *Pervasive Computing Technologies for Healthcare, 7th International Conference on Pervasive Health*, pp. 17-24, IEEE.

Sharma, N. and Gedeon, T.D. (2012) Objective measures, sensors and computational techniques for stress recognition and classification: A survey, *Computer Methods and Programs in Biomedicine*, vol. 108:3, pp. 1287-1301. doi: 10.1016/j.cmpb.2012.07.003.

Zhu, D., Gedeon, T. D. and Taylor, K. (2011) "Moving to the Centre": A gaze-driven remote camera control for teleoperation, *Interacting with Computers (IwC)*, Vol 23:1, pp. 85-95.

Vo, T., Gedeon, T.D. and Mendis, B.S.U. (2010) Gaze pattern and reading comprehension, *ICONIP2010, LNCS 6444*, pp. 123-130.

Gedeon, T.D., Zhu, D. and Mendis, B.S.U. (2008) Eye gaze assistance for a game-like interactive task, *International Journal of Computer Games Technology*, doi:10.1155/2008/623725, pp. 1-10.

Zhu, D., Mendis, B. S. U., Gedeon, T. D., Asthana, A. and Goecke, R. (2008) A hybrid fuzzy approach for human eye gaze pattern recognition, Proceedings of 15th International Conference on Neural Information Processing (ICONIP'08), pp. 655-662.

Gedeon, T.D. (1997) Data mining of inputs: Analysing magnitude and functional measures, *International Journal of Neural Systems*, vol. 8 (2): 209-218.

Gedeon, T.D., Kozek, K. (1997) Neural Network and Hidden Markov Model Classification of Schizophrenia using Eye Gaze Data, *Proceedings 2nd International ICSC Symposium on Soft Computing (SOCO'97)*, 382-385, Nimes.

Treadgold, N.K. and Gedeon, T.D. (1997) A Cascade Network Algorithm Employing Progressive RPROP, in Mira, J, Moreno-Díaz, R and Cabestany, J, (eds.), *Biological and Artificial Computation: From Neuroscience to Technology*, 733-742, Springer Verlag, Lecture Notes in Computer Science, vol. 1240.

Please note that some of these papers were mentioned in asides or in response to questions and so may not be directly referred to in the text of the slides. I am happy to reply to questions as to their relevance. ☺