Jaana Simola, Jarkko Salojärvi, and Ilpo Kojo. 2008. Using hidden Markov model to uncover processing states from eye movements in information search tasks. Cognitive Systems Research, volume 9, number 4, pages 237-251.

# Using Hidden Markov Model to uncover processing states from eye movements in information search tasks

Jaana Simola *

*Humanities Lab, Centre for language and literature,*
*Lund University,*
*S-22100 Lund, Sweden*

Jarkko Salojärvi

*Adaptive Informatics Research Centre,*
*Department of Information and Computer Science,*
*Helsinki University of Technology,*
*P.O.Box 5400, FI-02015 TKK, Finland*

Ilpo Kojo

*Center for Knowledge and Innovation Research,*
*Helsinki School of Economics,*
*P.O.Box 1210, FI-00101 Helsinki, Finland*

## Abstract

We study how processing states alternate during information search tasks. Inference is carried out with a discriminative hidden Markov model (dHMM) learned from eye movement data, measured in an experiment consisting of three task types: (i) simple word search, (ii) finding a sentence that answers a question and (iii) choosing a subjectively most interesting title from a list of ten titles. The results show that eye movements contain necessary information for determining the task type. After training, the dHMM predicted the task for test data with 60.2% accuracy (pure chance 33.3%). Word search and subjective interest conditions were easier to predict than the question–answer condition. The dHMM that best fitted our data segmented each task type into three hidden states. The three processing states were identified by comparing the parameters of the dHMM states to literature on eye movement research. A *scanning* type of eye behavior was observed in the beginning of the tasks. Next, participants tended to shift to states reflecting *reading* type of eye movements, and finally they ended the tasks in states which we termed as the *decision* states.

*Key words:* Eye movements, Computational models, Hidden Markov model, Information search, Scanning, Reading, Decision process

* Jaana Simola
*Email addresses:* jaana.simola@helsinki.fi (Jaana Simola), jarkko.salojarvi@tkk.fi (Jarkko Salojärvi), kojo@hse.fi (Ilpo Kojo).

## 1 Introduction

Eye movements are commonly used as indicators of on-line reading processes because of their sensitivity to word characteristics. Empirical evidence supports this eye-mind link assumption: longer eye fixations have been observed together with misspelled words, less common words, or words that are unpredictable from their contexts (Rayner, 1998; Rayner and Pollatsek, 1989). However, reading studies typically concentrate on microprocesses of reading, such as studying how word features determine when and where the eyes move. Moreover, their analysis of eye movement data is often based on linear models that fail to consider eye movements as time-series data and therefore do not account for variations within a task.

Our contribution is to analyze the whole sequence of fixations and saccadic eye movements to gain an insight into how processing alternates during the reading task. In other words, we assume the reverse inference approach, and try to infer the hidden cognitive states from an observable eye movement behavior (see Poldrack (2006) for a discussion on the possible benefits and pitfalls of the approach within neuroimaging research). The relationship between eye movements and cognitive states is modeled with a discriminative hidden Markov model (dHMM). In our application, we use the dHMM to map the changes in statistical patterns of eye movements to changes of the hidden states of the model as participants proceed in information search tasks. A hypothesis on the cognitive states corresponding to the hidden states can then be made by comparing the parameters of the hidden states (for example fixation durations and saccade lengths) to literature on eye movement research where the cognitive state is known.

The states discovered by our model suggest that processing alternates along the completion of the tasks, even when the abstractness of the searched topics varies. The results can be used in practical applications. Earlier, Hyrskykari et al. (2000, 2003) have used the fact that fixations are longer during processing difficulties in order to develop an interactive dictionary that gives translation aid when it detects reading difficulties. However, detecting changes in processing states makes it possible to develop more advanced applications. For example, a proactive information retrieval application can search for more documents on a specific topic after detecting eye movements that indicate careful processing when a person is reading about that topic (see Puolamäki et al. (2005) for a feasibility study). The goal of the present article is to show that prerequisites for implementing such techniques exist.

Previously, Carver (1990) has argued that readers use different processes in order to better accomplish their goals. They change their ongoing process either by instructions or by the difficulty of the text. Carver distinguishes five basic processes based on variations in reading rates, that is, the number of words covered by reading time (i.e. words per minute, wpm). The suggested processes are called *scanning, skimming, 'rauding', learning* and *memorizing. Scanning* is performed at 600 wpm and is used while the reader is searching for a particular word in a text. Another rapid and selective process is *skimming* (450 wpm), which is used in situations where the reader tries to get an overview of the content without reading through the entire text. *'Rauding'* (300 wpm) corresponds to normal reading in which the reader is looking at each consecutive word of a text to comprehend the content. *Learning* is slow (200 wpm) and is used for knowledge acquisition. *Memorizing* is the slowest process (138 wpm) and involves continuous checks to determine whether the ideas encountered might be remembered later.

According to Carver, the processes represent different cognitive processes and he suggests that readers shift between them, in a manner similar to drivers shifting gears. He also suggests that skilled readers vary their reading processes more than poor readers. The eye movement results indicate that when participants switched up, for example, from the *'rauding'* to the *skimming* process, the mean fixation durations decreased together with the mean number of fixations and regressions (i.e. fixations back to previously read text). Also the length of forward saccades increases. On the other hand, switching down resulted in more regressions, longer fixation durations, and shorter saccade lengths.

Carver suggests that the primary factor influencing reading rate is the selected reading process. Minor within-process variations result from the difficulty of the text and individual differences, such as age, practice or cognitive speed. Previous research indicates also between-individual differences in reading strategies (Hyönä et al., 2002).

### *1.1 Models of eye movement control during reading*

Computational models on eye movement control during reading have been successful in explaining how various perceptual, cognitive and motor processes determine when and where saccades are initiated during reading. The current controversy is whether attention in reading is allocated serially to one word at a time, as suggested by the E-Z Reader model (Reichle et al., 2006; Pollatsek et al., 2006), or whether attention is spatially distributed so that several words are processed at the same time. This parallel hypothesis is supported for example by the SWIFT (Richter et al., 2006), the Glenmore (Reilly and Radach, 2006) and the Competition/Interaction (Yang, 2006) models. (For a review of the computational models of reading, see: Cognitive Systems Research, 2006, 7, pp.1-96.) However, these models are limited in their ability to consider variations in higher level reading processes.

The models mentioned above construct very specific hypotheses on the reading process and thus use tailored parameter values developed in accordance with what is previously known about human vision, such as the size of the visual span and variability in saccade and fixation metrics, as well as word recognition processes like the time for lexical access. Instead of fixing model parameters manually, the model parameters can also be learned from the data. The general idea is that information required for constructing a model is learned from the empirical data, for example the best model structure or the best parameter values. To avoid overfitting, the data is split into two subsets: training and testing data sets (see e.g. Hastie et al. (2001)). The best model and its parameters are selected using the training data, and then its generalization capability (i.e. how well the model fits new data) is tested using the test data. Feng (2006) has applied similar approach for modeling age-related differences in reading eye movements.

### *1.2 Purpose of the study*

Our goal is to investigate how processing changes as the participants proceed in three types of information search tasks: simple word search, question–answer task and finding subjectively most interesting topic. For this purpose, we combine experimentation with data-driven modeling using a discriminative hidden Markov model (dHMM). As a time-series model it is well suited for our purposes because it provides a more comprehensive description of the eye movement pattern than the basic summary statistics such as average fixation duration. To capture the relationship between language processing and eye movements, we model the observed time series of fixations and saccades by assuming latent states that are supposed to be indicators of the cognitive system that switches between different states of processing. We assume that in each processing state the statistical properties of the eye movement patterns are different. The best model topology, that is, the number of hidden states, is found by comparing several possible model topologies with cross-validation, and choosing the one that best explains unseen data. We also compare the parameter values of the model to what is previously known about reading and performance in other cognitive tasks. This information is used to make inference about processing during the tasks.

Our approach is not committed to any particular processing theory. Therefore many of the theoretical issues discussed in eye movement models of reading (Pollatsek et al., 2006; Reichle et al., 2006; Reilly and Radach, 2006; Richter et al., 2006), such as the parafoveal-on-foveal effects, do not concern our model. Instead, the dHMM applied here describes how eye movement behavior varies during a single trial, and the states uncovered by the dHMM can be seen as hypotheses about the ongoing processes which are based on the statistical regularities of the eye movement data.

## 2 Data collection

### *2.1 Participants*

Eye movement data were collected from ten volunteers (6 female). The age range was 23–29 years, mean age 25.7 years, (SD = 1.9). They had normal or corrected to normal vision and all of them were native speakers of Finnish. Participants filled in a written consent before the experiment.

### *2.2 Procedure*

Our tasks represented single online information search episodes where the user is inspecting listings returned by a search engine in order to find a topic of her interest. The task types were selected to fit the possible practical implementation, a proactive information retrieval application. The task of the participants was to find a target from a list of ten titles. The level of complexity in the searched topics varied through the inclusion of three different types of tasks:

1. **Word search (W):** The task is to find a word from the list.
2. **Question-answer (A):** A question is presented and the task is to find an answer to the question from the list.
3. **True interest (I):** The participants are instructed to search for the most interesting title in the list.

The trial structure was similar across the tasks (Figure 1). First, the assignment was presented: The participants saw a sentence instructing them to find either a word (W), an answer to a question (A), or the most interesting sentence (I), according to the condition. After the assignment, a list of sentences was presented. The participants were instructed to view the list until they had found the relevant line. Eye movements were recorded during this period. After finding the relevant line, they pressed 'enter', and were shown the same sentences with line numbers. They then typed the number corresponding to the line they had chosen. Before the experiment, participants read the instructions and practiced each of the tasks.

Each participant conducted a total of 150 assignments. The experiment was divided into 10 blocks, with 15 assignments in each block. Each task type was presented five times within a block. The presentation order of the blocks and the assignments within them was randomized.

(Figure 1 about here)

### *2.3 Stimulus material*

The text material consisted of 500 online newspaper titles, revised to grammatical sentences. The maximum length of the sentences was 80 characters. On average, there were 5.8 words per sentence and the mean word length was 9.9 characters. The sentences were divided to 50 lists of ten sentences. To control for the effects of previous topic knowledge, the sentences were selected to represent three general topics: Finnish homeland news (20 trials), foreign news (20 trials) and business & finance news (10 trials). The texts were written in Finnish, and a 30-point Arial font was used. The average character height was 0.9 degrees and the average character width was 0.5 degrees from the viewing distance of about 60 cm.

For the word search condition, fifty words were chosen as target words. The positions of the targets in sentences were balanced, i.e., the words appeared equally often as the first, second, third or fourth word of the sentences. For the question-answer condition, we prepared fifty questions, which were validated with a pilot test including eight participants. We modified the questions and sentences until their answers agreed in 74 % of the trials, and conducted the actual experiments with the modified questions and sentences. In word search and question-answer conditions, the locations of the correct lines were balanced so that the answers appeared equally often in all ten sentence-lines. For the true interest condition, no additional stimulus preparations were needed.

To emphasize the differences between tasks and to minimize stimulus-driven factors on processing, the same stimuli were presented in all three task types. In order to control for the possible effects of repetition, a set of analysis was carried out with repeated measures ANOVAs. We found no significant effect of presenting the same stimulus three times during the experiment on the number of fixations (F(2,18) = 2.86, *ns.*), average fixation durations (F(2,18) = .18, *ns.*) or saccade lengths (F(2,18) = 1.00, *ns.*) in an assignment. Therefore we did not have to consider the effect of stimulus repetition in our modeling work.

### *2.4 Apparatus*

The stimuli were presented on a 17 inch TFT display with a screen resolution of 1280 x 1024 pixels. The display was located on a table at the eye level of the participants, at the distance of approximately 60 cm. In order to maintain the life-likeness of our setup, no chin or forehead rests were used for stabilizing the heads of the participants.

Eye movements were recorded by a Tobii 1750 remote eye-tracking system with a spatial accuracy of 0.5 degrees. The screen coordinates of both eyes were collected from each participant at 50 Hz sampling rate. The eye tracking system was calibrated between the experimental blocks using a set of sixteen calibration points shown one at a time.

### *2.5 Preprocessing*

Fixations were computed from the data using a window-based algorithm by Tobii. Visualizations of measured gaze coordinates were used to choose fixation window parameters for further analysis. Based on the visual inspections we selected three candidate parameter setups: (i) a 20 pixel widow with a minimum fixation duration of 40 ms, (ii) a 40 pixel window with 80 ms fixation duration, and (iii) a 20 pixel window with 100 ms fixation duration. Blinks were left out from the raw data by the Tobii software, otherwise no editing of the eye movement data was carried out.

The best fixation window parameters were determined using the logistic regression model (see Sections 3.1 and 3.4.1) and a 40-fold cross-validation (see Section 3.5) of the data. The procedure produced 40 perplexity values for left-out data with each of the fixation window parameter combinations.

For the Tobii 1750 eye tracker, the fixation window that resulted in best classification accuracy ($p < .05$, Wilcoxon signed rank test) of the left-out data sets was a 40 pixel window of 80 milliseconds (3.2 letter spaces).

## 3 Modeling

The total data consisted of 1456 eye movement trajectories, that is, fixation-saccade sequences measured from each assignment. 44 trials were missing because no eye movements were measured, for example due to double key pressings of the subjects. The total data were randomly split into a training set of 971 trajectories and a test set of 485 trajectories.

Throughout the analysis we used a data-driven approach: the data was used for making decisions on different modeling questions. Best model topology was selected by using cross-validation with the training data. Parameters of the best model were then learned using the full training data, and the generalization capability, i.e., how well the model fits unseen data, was tested with the test set. The reason for using test data is that with increasing model complexity, that is, with increasing number of parameters, the model will more accurately fit the training data. At some point this turns into *overfitting*, where increasing the model complexity will decrease the model performance on unseen data whereas the performance on training data set continues to increase.

### *3.1 Logistic regression*

In our experiment the ground truth for a given eye movement trajectory, that is, the information about the task type, was always available. Suitable models for such data belong to the general category of supervised or *discriminative* models. The simplest discriminative model is logistic regression (see Hastie et al. (2001)), which predicts the probability of class (task type), conditional on covariates (the associated measurement data) and parameters. The covariates are assumed to be given, that is, no uncertainty is associated with their values. The model is optimized by maximizing the conditional likelihood. However, logistic regression cannot model time series data. A common approach is to compute some form of statistics from the time series and then use these as covariates.

We used logistic regression as a simple classifier to obtain baseline results for the HMM, and for selecting the best fixation window parameters.

### *3.2 Hidden Markov Models*

To analyze the fixation-saccade sequence as a time series we used Hidden Markov model, which is commonly used for analyzing sequential data, such as speech (see e.g. Rabiner (1989) for an introduction on HMMs). The HMMs belong to the general category of *generative joint density* models which attempt to describe the full process of how the data is being created, that is, they do not use covariates. Whereas fully discriminative models concentrate only on separating different classes, and thus provide no physical interpretation of the parameter values, the parameters of a joint density model can be associated with the data, giving an insight into the underlying process, assuming that the model describes the data accurately enough. HMMs are optimized by maximizing the log-likelihood, $\log p(C, X|\Theta)$, of the data $C \cup X$, given the model and its parameters $\Theta$. Here $X$ is the observation sequence, eye movement trajectory, associated with class $C$, the task type.

HMMs are applied in a case where the statistical properties of the signal change over time. The model explains these changes by a switching of a hidden (unobservable, latent) state $s$ within the model. The total number $S$ of hidden states can be learned from data, for example by cross-validation. Each of the states addresses an associated observation distribution $p(\mathbf{x}|\theta_s)$, from which the data is generated. The parameters $\theta_s$ can be different for each state (e.g. for Gaussian distributions having different means and standard deviations). The changes in the distributions of the observations are thus associated with transitions between hidden states. The transitions are probabilistic, and defined by a transition matrix $\mathbf{B}$. We assume a first-order Markov property for the transitions, that is, we assume probabilities $p(s(t+1)|s(t))$; the transition to the next state $s(t+1)$ depends only on the current state $s(t)$. Pieters et al. (1999) showed that eye movements follow this property. Additionally, this restricts the number of parameters in the model, making modeling computationally more efficient.

A full definition of HMMs requires one more set of parameters, $\pi(s),\ s = 1 \dots S$, which is the probability of initiating the time sequence at state $s$. An example topology of an HMM is illustrated in Figure 2.

For a time series $\mathbf{x}_{1 \dots T}$ of observations the full likelihood of the HMM is then

$$p(\mathbf{x}_{1 \dots T}|\Theta) = \sum_{\mathcal{S}} \pi(s(1)) p(\mathbf{x}(1)|s(1)) \prod_{t=2}^{T} p(\mathbf{x}(t)|s(t)) p(s(t)|s(t-1)), \tag{1}$$

where $\mathcal{S}$ denotes all "paths" through the model, that is, all $S^T$ combinations of hidden states for a sequence of length $T$, and $\mathbf{x}(t)$ is the measured observation vector at time $t$.

Maximum likelihood parameter values of the HMMs are obtained with the Baum-Welch (BW) algorithm, a special case of Expectation-Maximization (EM) algorithm, which can be proven to converge to a local optimum. Fast computation of the most probable path (hidden state sequence) through the model, given a new data sequence, is obtained using the Viterbi algorithm.

Previously, Liechty et al. (2003) applied hidden Markov models to study two states of covert attention, local and global attention. They showed that viewers were switching between the attention states while they were exploring print advertisements in magazines. The local visual attention state was characterized by short saccades, whereas in the global attention state, longer saccades were common. In another line of research, Salojärvi et al. (2005b) showed that perceived relevance of a text could be predicted from eye movements in an information search task.

### *3.3 Discriminative Hidden Markov Models*

A generative model can be converted to a discriminative model by optimizing the conditional likelihood of the model $\log p(C|X, \Theta)$, obtained from a generative model via Bayes formula. Compared to a fully discriminative model (such as logistic regression), the converted model still has the benefits of a generative model, such as easier interpretation of model parameters (see Salojärvi et al. (2005c) for a description of the differences).

Discriminative training of HMMs is carried out by assigning a set of "correct" hidden states $\mathcal{S}_c$ in the model to always correspond to a certain class $c$, and then maximizing the likelihood of the state sequences that go through the "correct" states for the training data, versus all the other possible state sequences $\mathcal{S}$ in the model (Povey et al., 2003; Schlüter and Macherey, 1998).

The parameters of a discriminative HMM (dHMM) are optimized with a discriminative EM (DEM) algorithm, which is a modification of the original BW algorithm (the derivation of the algorithm can be found in Salojärvi et al. (2005a)).

## *3.4 Feature extraction*

### *3.4.1 Features for logistic regression model*

The logistic regression was used as a baseline to a HMM. It uses averaged features that can be derived from the fixation-saccade time sequence, i.e., it obtains the same information as the HMM. The features were:

(1) Length of the sequence (number of fixations).
(2) Mean of fixation duration (in milliseconds).
(3) Standard deviation of fixation duration.
(4) Mean of saccade length (in pixels).
(5) Standard deviation of saccade length.

### *3.4.2 Features for hidden Markov model*

For the time series model, four features of each fixation were computed from the eye movement trajectory, that is, from the raw fixation-saccade data from each assignment. The features are listed below with the corresponding modeling distribution (the distributions denoted by $p(\mathbf{x}|s)$ in Equation (1)) reported in parenthesis. See e.g. Gelman et al. (2003) for the parametric form of the distributions.

(1) Logarithm of fixation duration in milliseconds (one-dimensional Gaus-

sian).

(2) Logarithm of outgoing saccade length in pixels (one-dimensional Gaussian).

(3) Outgoing saccade direction (quantized to 4 different directions) + a fifth state indicating that the trial had ended (Multinomial).

(4) Indicator variable of whether there have been previous fixations on the word which is currently fixated (Binomial).

In literature (e.g. Reichle et al. (2006)), a gamma distribution has often been used for modeling fixation durations, because its negatively skewed distribution resembles the data. There are two alternatives to implement this. In the first version, the data sequence is indexed by time, and thus the hidden state sequences are directly mapped into fixation durations (Liechty et al., 2003), and therefore the probability of staying in state $s$ must follow a gamma distribution. However, in ordinary HMMs this probability follows an exponential rather than gamma distribution, and therefore a semi-hidden Markov model needs to be implemented, where the transition probabilities depend on the time spent in the current hidden state. We here applied the second alternative. We constructed a HMM that emitted the fixation durations, changing the time scale of the HMM into fixation counts. Instead of having a HMM that is in state $s$ for the time $t \ldots t + \tau$, we now have a HMM that is in state $s$ for fixation $i$, which has the duration $\tau$. We then make a simplifying assumption by modeling the logarithm of fixation durations with a Gaussian. Further work could include extending this model to a mixture of two log-normal distributions, since this has been found to work well for reading fixations (Carpenter and McDonald, 2007).

The saccade lengths were quantified as pixels and were not converted to more conventional measures, such as characters or degrees during computations, because conversions would have added noise to data (since the Tobii 1750 allows free head movement). Saccade lengths were computed from the raw 50 Hz gaze data by computing the distance between the gaze location at the end of the previous fixation and the beginning of the current fixation. The spatial accuracy of the eye-tracker was 0.5 degrees corresponding to approximately 12 pixels.

For saccade quantization, each fixation was first mapped to the closest word in the preprocessing stage. The outgoing saccade direction was then encoded with an indicator variable that can obtain five different values: 1 – saccade forward on the current line of text, 2 – saccade upwards from the current line, 3 – saccade backwards on the current line, 4 – saccade downwards from the current line, and 5 – ending the assignment.

### *3.5 Model selection*

When choosing fixation window parameters or the number of hidden states of the HMM, an $n$-fold cross-validation with the training data was carried out. In this procedure, the training set is divided into $n$ non-overlapping subsets, and each of the subsets is in turn left out as a validation data set. The training is carried out using the other $n - 1$ subsets, and then the generalization capability of the model is tested with the validation set. The procedure is carried out for all alternative modeling configurations. The method produces $n$ paired measures of goodness of model fit, calculated from validation data, allowing us to test the out-of-sample performance of the model configurations.

The reason for using cross-validation is to avoid overfitting, i.e., choosing a too complex model. Alternative methods for model selection include a computationally much heavier bootstrap method (Efron and Tibshirani, 1993), or using information theoretic criteria (Akaike, 1974; Schwartz, 1978). The latter however are not theoretically justified in case of HMMs, see for example Robertson et al. (2004), and the references therein.

Goodness of the model was measured in two ways; in terms of classification accuracy and perplexity. Classification accuracy is the amount of correctly predicted task types divided by the total amount of tasks. However, for relatively small data sets, the classification accuracy is a noisy measure, since each sample can be assigned to only one class. A better measure is therefore the perplexity of the test data set, which measures the confidence in the predictions of the classifier. It is defined as a function of the average of log-likelihoods $\mathcal{L}$ of the $N_s$ test data sequences, denoted formally by

$$\text{perp} = e^{-\frac{1}{N}_s \sum_{i=1}^{N_s} \mathcal{L}_i}; \quad \mathcal{L}_i = \log p(c_i | x^i_{1\ldots T_i}, \theta) \quad , \tag{2}$$

where $x^i_{1\ldots T_i}$ denotes the $i$th sequence of observations of length $T_i$, and $c_i$ is the type of task $i$. $N_s$ is the number of sequences, and $\theta$ the model parameters. The best possible perplexity is 1, where the correct task type is predicted with a probability 1. On the other hand, perplexity of 3 corresponds to random guessing with a probability of $\frac{1}{3}$ for each of the task types. In our data analysis, the class distribution was not equal within the training and test sets. This was mainly due to random split of the data, and in part due to missing eye movement measurements. If these are taken into account, the random perplexity for the test set is 3.01. If perplexity is greater than this the model is doing worse than random guessing. In the worst case where the classifier gives a (close to) zero probability for the correct class, the perplexity is restricted to a maximum value of $10^{22}$.

## 4 Results

### *4.1 Logistic regression*

The results of the logistic regression are reported in Table 1. The perplexity of the test set was 2.42 with a classification accuracy of 59.8 %.

(Table 1 about here)

### *4.2 Discriminative hidden Markov model*

All modeling with HMMs was carried out in a data-driven fashion. The topology of a HMM was fully connected, that is, transitions between all states were possible. All parameter values were learned from data by maximizing the conditional likelihood. The number of hidden states in the dHMM was determined with a 6-fold cross-validation. The different hidden state configurations that were tried out were $S \in \{$2-2-2,2-2-3,2-3-3,3-3-3,3-3-4,3-4-4,4-4-4$\}$, corresponding to the number of hidden states used for modeling word search, question-answer and true interest conditions, respectively. The scheme for increasing the number of hidden states in the HMM was arrived at after observing that the eye movement trajectories were usually longest in the true interest condition and then in the question-answer condition.

The number of hidden states was decided as in Robertson et al. (2004) by comparing the mean of perplexities of validation sets. The decrease of out-of-sample perplexities started to level off when the number of hidden states was nine, suggesting that this is the optimal number of hidden states. Since the variance of conditional maximum likelihood estimates is larger than maximum likelihood estimates (Nádas, 1983), we additionally compared the paired perplexity values for eight, nine, and ten hidden state configurations with a Wilcoxon signed rank test. The difference between the 8-state and 9-state models was statistically significant ($p = 0.03$), whereas the difference between 9-state and 10-state models was not. Since the data does not support the preference of a 10-state model over a 9-state model, the less complex model should be preferred. The model with nine hidden states is obtained also when using a majority vote-based model selection scheme (Miloslavsky and van der Laan, 2002).

The 9-state HMM achieved the perplexity 2.32 and classification accuracy of 60.2 % for the test data. The confusion matrix of the dHMM is reported in Table 2. Both logistic regression and dHMM could separate the two extremes, word-search and true interest, but predicting the question-answer -tasks is difficult. One possible reason is that some of the question-answer assignments were easier than others. The search behavior in easy assignments may have resembled the fixation patterns in word search task (in case where the question can be answered with one word), whereas difficult question-answer assignments were confused with the task of indicating subjective interest.

### *4.2.1 Comparing the classification accuracies and perplexities*

If the time series of the eye movement data contains information about the task type, the dHMM should perform better than logistic regression model using averaged features. The perplexity of the test set for dHMMs was 2.32, whereas logistic regression achieved the perplexity of 2.42. The dHMM was significantly better than logistic regression ($p < .01$, comparison of perplexities with a Wilcoxon signed rank test). The time series of the eye movements therefore contained relevant information for determining the task type.

(Table 2 about here)

## *4.3 Interpreting HMM parameters*

Proper interpretation of the parameters of a discriminatively trained joint density model (e.g., a dHMM) is still a somewhat open question. Based on asymptotic analysis (with infinite data), following can be said.

Ordinary maximum likelihood training of a joint density model minimizes the Kullback-Leibler divergence (Cover and Thomas, 1991) between the data and the model parameters. This can be seen by considering the data to be generated from a "true", however unknown, model with model parameters $\tilde{\theta}$. In practise the model is always an approximation of the "truth", and therefore the model will not fit perfectly to the data (if it were perfect, it should predict all unseen data perfectly) This incorrectness causes a bias in the obtained model parameters $\theta$.

Discriminative training, on the other hand, maximizes conditional likelihood which minimizes the Kullback-Leibler divergence between a *subset* of variables in the data and the model parameters. As a result, this subset (here the task types) is modeled as well as possible. A tradeoff is that other variables of the data are modeled more inaccurately. However, in an asymptotic case with infinite amount of data, and where the "true" model is within our model family, the parameters are the same as those obtained from maximum likelihood. In case of an incorrect model, by inspecting the gradient of the conditional likelihood (proof omitted), it can be shown that the conditional maximum likelihood and the maximum likelihood estimates are close to each other (and asymptotically the same) when (i) the model is close to the true model, or (ii) the class predictions of the model are accurate, but the particular parameters do not help in discriminating between the classes. In these cases the parameters can be interpreted as in an ordinary joint likelihood model.

From this point of view, a straightforward way of interpreting parameter values is therefore to report and compare the parameter values from conditional and ordinary maximum likelihood. If the values are same, the data does not contain additional information that can be used for more accurate prediction of the task type. On the other hand, if the two parameter estimates differ, it implies that the variables that they model help in predicting the task type, and their modeling assumptions are incorrect. This fact can be used for checking and revising the model. The revised model has to be checked afterwards with new data.

In our experiment the parameters of the discriminative and joint density HMMs (Table 3) are roughly the same, suggesting that our model uses the information that eye movements contain on task types fairly well. The greatest discrepancy between the parameter values follows from the log-Gaussian approximation of the fixation distributions, which was to be expected (as discussed in Section 3.4.2). The difference between the two parameter estimates also shows that the fixation durations are important in predicting the task type.

We next discuss modeling results of each set of parameters of HMM. Analysis is carried out with conditional maximum likelihood parameters; maximum likelihood parameters can be analysed in a similar manner, with approximately similar results.

(Table 3 about here)

### *4.3.1 Observation distributions and hidden states*

The discriminative hidden Markov model that best fitted our data segmented each task type into three states (Figure 2). The parameter values of the dHMM (Table 3) exhibited relatively similar eye behavior in the three hidden states for each of the task types. Next, we compared the parameter values to literature on reading and other cognitive tasks, and designated the states to describe the processing features that were reflected in the eye movement behavior.

(Figure 2 about here)

With a combined probability of 67 % (Table 3 and Figure 2), participants began the assignments from states which we termed as *scanning*, because the parameters suggested rather long saccades, with no clear preference on direction (i.e., almost random), and fewer saccades towards previously fixated areas. The fixation durations were relatively short (approximately 135 ms), which is in accordance with previous results indicating shorter fixations in association with easier tasks (Rayner, 1998). On average, participants spent 2.8 s scanning (Table 4).

The second set of states were labeled as *reading*, because they were characterized by frequent forward saccades (over 60 % probability) with an average fixation duration of about 200 ms, also typical for reading. The percentage of backward saccades was 12–15 %, corresponding to the previous findings suggesting that in normal reading about 10–15 % of saccades are regressions (Rayner, 1998). The average saccade length was 10.3–10.7 letters (128–133 pixels), which corresponds to the average length of a word (9.9 characters), plus a space between words.

Frequent forward and backward saccades were typical for the third and final states (Table 3). The percentage of backward fixations (20–30 %) was twice the amount usually observed in reading. Saccade lengths were approximately 10.7 letters (133 pixels), corresponding to the length of a word, and occurred within the same line (with 75 % probability). The fixations landed to previously fixated words with 78–86 % probability. On average, the fixation durations (175 ms) were shorter than in reading states. This is possibly due to the fact that participants were mostly fixating on words which they had recently seen, and therefore the lexical access took less time. We termed the third states as *decision* states, because the features indicated a lot of re-reading of the previously seen lines. Almost without exception, participants ended the assignments while they were in the third states. This pattern is visible in Figure 4. Shimojo et al. (2003) have reported similar results in the context of preference decisions made for faces. They also showed that participants tended to look more often at the target they chose just before they made their decisions.

One potential concern regarding the comparisons of parameters with previous reading studies, for example those reviewed by Rayner (1998), is that the participants may have varied their processing states also in the reviewed tasks. However, as brought out by Hyönä et al. (2002), in many reading studies, factors such as global reading strategies have been treated as a nuisance, and their influence is minimized by studying reading under simplified conditions (i.e. using brief and simple texts for very simple purposes). Therefore it is likely that previous results mostly reflect rather 'pure' types of processes.

### *4.3.2 Transition probabilities*

The transition probabilities of the dHMM are shown in Figure 2. Participants continued within the same processing state for several steps (i.e., fixations), indicating that the associated cognitive processes operate on time scales longer than one fixation. Similarly, previous research suggests that the on-going processes are not reset after every saccade, but their influence survives across saccades (Yang and McConkie, 2005). An estimate of these time scales was next obtained with the dHMM.

**4.3.2.1 Method.** The most probable state sequence for each eye movement trajectory was computed by applying Viterbi algorithm to the learned HMM. The means and standard deviations of the process durations (Table 4) were computed from the data using the state segmentation obtained from the dHMM. The mean is the average time spent in a state, and standard deviation describes how the time varies in individual cases. An error of the two estimates, i.e., how accurate the estimates are given in our (finite) data sample, is obtained with a bootstrap method (Efron and Tibshirani, 1993). We generate 400 replicate (bootstrap) data sets by sampling from the original data with replacement. For each of the replicate data sets a bootstrap estimate was computed (e.g. the mean). The error is now the standard deviation of the 400 bootstrap estimates computed with respect to the original estimate.

**4.3.2.2 Results.** Table 4 shows that the times spent in each of the states did not differ considerably across the task conditions. Participants spent more time in *scanning* and *reading* than in *decision* states. The decision times were two times longer for the question-answer and for the subjective interest conditions than for the word search, where the assignment was ended approximately 1 second after reaching *decision* state. This corresponds to the duration of making the decision, because the participants did not go back to *scanning* or *reading* states, unlike in other conditions. Also, the time to reach the *decision* state increased with the task complexity.

(Table 4 about here)

### *4.3.3 Transitions between states*

Figure 2 shows that in the word search condition, transitions from the *decision* state are rare, with only 1 % probability, whereas in the question-answer condition these transitions occur with 5 % probability and in the subjective interest condition with 14 % probability. In the word search and question-answer conditions, participants switched more often from *scanning* to *decision* (with 80 % probability) than to *reading* (20 % probability). This can be seen from Figure 2 by comparing the associated transition probabilities (8 % vs. 2 %). From *reading*, they shifted to the *decision* state. In word search, this probability was 92 % (11 % vs. 1 %), and in the question-answer condition 55 % (6 % vs. 5 %). In the true interest condition, there was a strong tendency to switch from *decision* to *reading* with 86 % probability (12 % vs. 2 %).

#### *4.3.4 Eye movement trajectories*

When combining the most probable (Viterbi) path through the hidden Markov model with the interpretations of the hidden states, it is possible to make hypotheses on the switches of the cognitive states during an assignment. An interesting further study would be to map these switches to text contents. Figure 3 shows example trajectories for the task types, plotted on the screen coordinates (stimulus words are not plotted for clarity). It appears that when the participant closes in to the relevant line, the *decision* state is adopted. In the word search condition, the trajectories indicate mostly scanning, whereas in question – answer condition the lines are read word by word, but the state of processing varies, depending on whether the line is relevant for the task or not.

(Figure 3 about here)

### *4.3.5 Average behavior*

Drawing summaries from the plots shown in Figure 3 is difficult. Instead, it is easier to find common patterns by inspecting the mean behavior of the conditions.

**4.3.5.1 Method.** Computing average behavior from our time series data is not straightforward, because time sequences have different lengths and the observations are probabilities. We first computed the a posteriori probabilities of being in state $s$ at time $t$, given the observations $\mathbf{x}_{1\ldots T}$ and model parameters $\theta$, that is, $\gamma_t(s) = p(s_t|\mathbf{x}_{1\ldots T}, \theta)$. The probabilities can be computed with a forward-backward algorithm. The probabilities were then converted to their natural parameters (by $\theta_{\gamma_t}(s) = \log \gamma_t(s)$, thus mapping the probabilities to real values). Next, the sequences were normalized to the same length by resampling them to the same length as the longest sequence (Gallinari, 1998). After that, the values were mapped back to probabilities using the inverse mapping $\gamma_t(s) = \frac{\exp\{\theta_{\gamma_t}(s)\}}{\sum_i \exp\{\theta_{\gamma_t}(i)\}}$. A simple assumption is that for each time instance $t$, the probabilities are emitted from a Dirichlet distribution with parameters $\alpha^{(t)}$. The parameters can be estimated using the maximum likelihood criteria (see Minka (2000) for update formulas), after which the mean and standard deviation of the Dirichlet distribution can be computed (see e.g. Gelman et al. (2003)).

**4.3.5.2 Results.** The mean behavior along with its standard deviation is plotted in Figure 4. In the word search condition, participants began the assignment from the *scanning* state with a probability of 70 %. There was a slight tendency for being in the *reading* before switching to the final *decision* state. For the question-answer and subjective interest conditions the strategies were similar, although they were less emphasized. Participants began the tasks almost equally often from the *scanning* and *reading* states. In the middle of the task performance, the *reading* state was slightly more common and towards the end, the *decision* state was very common. In general, the results suggested that before shifting to the *decision* states participants adopted different strategies. This was also visible in the standard deviations, which were larger in the beginning and in the middle of the tasks than in the end.

(Figure 4 about here)

## 5 Discussion

In this paper, we applied a reverse inference approach with the aim of making hypotheses on hidden cognitive states in an experiment resembling everyday information search tasks. Our setup differs from traditional research methods in psychology where controlled experiments are designed to find out what happens in eye movements when cognitive processes are manipulated. Instead, we designed a less controlled experiment, and then applied advanced statistical modeling, a hidden Markov model to make inferences about cognitive processing during the tasks (see Feng (2003) for a discussion on benefits of the data-driven approach).

Our model suggests that participants shifted their eye movement behavior while they proceeded in tasks. They typically began the assignments from a set of states reflecting a *scanning* type of behavior (see Figure 4 and Table 3).

The scan paths indicated long saccades with no preference on direction, accompanied with rather short fixations. Additionally, the fixations tended to land on previously unfixated areas on the text.

The second set of states were labeled as *reading* because they contained frequent forward saccades, and the distance covered by saccades mostly corresponded to an average word length. Also the mean fixation durations (200 ms) and the amount of regressions (about 13 %) were in accordance with the previous research findings of reading (Rayner, 1998).

The characteristics of the third set of states suggested a more careful analysis of sentences, possibly of *deciding* whether the sentence is the correct answer to a given task. This was indicated by the fact that the participants ended the assignments while they were in the *decision* states. The saccades landed almost always on the previously seen lines and were directed either forward or backward. The distance covered by saccades was about the length of an average word.

Our results support and complement the modeling work by Liechty et al. (2003), who used eye movement data to identify two states of visual attention in an advertisement viewing task. As an extension to their approach our model includes experimental manipulations of the search tasks. Although we used literal tasks, our processing states shared similarities with their findings. The *scanning* state had similar features with their global processing state, which were both characterized by long saccades and rather short fixations. Short saccades and long fixations were typical of their attentive processing state. In our study, the empirical data supported segmenting the attentive state into two processes, i.e. the *reading* and the *decision* processes, suggesting a finer structure.

Besides their behavioral relevance, the labels given to the hidden states are suggestive, and can be used as hypotheses about the underlying processes. The hypotheses can be tested by collecting additional data with known processing states, for example by selecting tasks that emphasize pure visual scanning or naturalistic reading, to empirically validate the parameters of suspected processes. With the setup presented here, it is also possible to make more specific hypotheses by constraining the dHMM structure. For example, some of the overlapping processes across the three tasks could have been linked in the HMM training.

For mutually exclusive processes the probability for being in one state at a certain time would be either one or zero. However, the probabilities suggested by our model were somewhere between one and zero (see Figure 2), indicating that the states are not mutually exclusive but rather reflect mixtures of ongoing processes that are optimal for the performance. This is in accordance with an experimental and theoretical evidence suggesting that reading eye movements are generated through multiple competing processes rather than one homogenous mechanism (Findlay and Walker, 1999). In addition, a considerable proportion of variation in eye movements can be attributed to random fluctuations in the oculomotor system (Feng, 2006). Also, McConkie and Yang (2003); Yang and McConkie (2005) have shown that a considerable amount (even 50 %) of saccades during reading are executed by a basic mechanism that repetitively produces saccades without direct cognitive control.

Our model was able to predict the task types with an accuracy of 60.2 %, which is 27 percent units above pure chance (33.3 % for three classes). We did not expect much better accuracy. First, because we used all data in modeling, including participants with noisier eye movement signals. Second, the tasks were not very controlled, instead the instructions allowed participants to freely choose their own search strategies. Third, the 50 Hz sampling rate of the Tobii 1750 eye tracker quantized the fixation durations to 20 millisecond intervals. With a higher temporal resolution the model may have been able to predict the tasks more accurately, since more information would have been available. The classification accuracy could also be improved by giving word level features, such as word frequencies and word lengths as an input to the model. This feature can be implemented for example by using a IOHMM model (Bengio, 1996; Bengio and Frasconi, 1999). Currently, the only additional information (besides eye movement data) given to our model was the task type of the learning data. Despite the moderate classification accuracy, the model parameters appeared behaviorally relevant when compared to the previous results about reading.

### *5.1 Relation to other models*

The model applied here, dHMM, makes it possible to study cognitive control across fixations, since the eye movements are inspected as a time series instead of summary measures, such as average fixation duration. Since the HMM is a model designed for reverse inference tasks, it differs from traditional computational models in psychology that are models of forward inference; they attempt to describe how perceptual and cognitive processes drive eye movements, whereas our model tries to make conclusions about cognition given the eye movements.

According to the visuo-oculomotor research tradition, non-cognitive factors, such as the landing position of the eyes on a word, mainly determine when and where the eyes move. Furthermore, Vitu et al. (1995) showed that eye movements varied little from normal reading when participants were pretending to read z-strings (however see Rayner and Fischer (1996)). Similar results were also shown by McConkie and Yang (2003); Yang and McConkie (2005). A strategy-tactics model  (O'Regan, 1990, 1992) suggests that, based on their expectations about the difficulty of the forthcoming task, readers can adopt either careful or risky global strategies that coarsely influence fixation times and saccade lengths. He claims that predetermined oculomotor strategies are important in defining global characteristics of eye movement behavior in reading. In our tasks, the question presented prior to the sentence lists most probably primes expectations and adjusts certain strategies for the forthcoming performance. Also, the states discovered by dHMM showed similar features across the task types. Therefore, it is possible that an oculomotor strategy optimized for the given tasks could explain the variations in processing states.

Other theories have emphasized the role of cognitive control on eye movements. For example, Just and Carpenter (1980) have proposed that eye movements act as direct pointers indicating which word is being processed and for how long. Also, computational models on reading eye movements, such as the E-Z Reader (Reichle et al., 2006; Pollatsek et al., 2006), are based on the assumption that fixation durations, word skipping or regressing are determined by lexical processes. However, the current discussions on the cognitive control theory focus on the decisions of when and where the next saccade is initiated within a single fixation. In contrast, the strategic control across fixations is until recently treated marginally.

In our tasks, the participants could have adjusted their processing states on moment-to-moment basis according to the current task demands, as proposed in Carver (1990). The finding that the task types differed in the transition sequences between the processing states could support the cognitive control theory. For example, in the question – answer and the subjective interest conditions, participants switched more often from the decision state back to the reading state, whereas in the word search condition the sequence was more straightforward, starting from the scanning state and ending in decision state.

### *5.2 Future directions*

As discussed above, both cognitive and oculomotor theories can explain our results. Therefore further studies, for example combining fMRI and eye tracking, could provide valuable information about the activities that correlate with the processing states reflected in eye movement patterns. For instance, emphasized simultaneous activation in language areas could support the cognitive control theory, whereas stronger correlations with motor activities would indicate that the strategies are determined by oculomotor factors.

In spite of the controversial views about the basis of the processes driving eye movements, our results are useful in practical applications. The finding that eye movement patterns differ when different processing demands are encountered can be used for developing an interactive information search application that learns and adapts to users' goals and intentions. For example, by examining which parts of a search engine results are read in different states, such as *reading* or *decision* states, it is possible to infer about the intentions and interests of the user. On the basis of this information the system could provide more material which is of possible interest to her. However, further studies are needed to make this kind of proactivity from the side of the system most beneficial to the users.

For future research more detailed experiments need to be designed, allowing deeper examination of the findings presented here. For example, it would be of interest to study to what extent the processing states generalize to other cognitive tasks and how individuals differ in switching between processing states.

### Acknowledgements

This work was supported by the Academy of Finland, decisions no. 202211 and 202209, Helsingin Sanomain 100-vuotissäätiö, NordForsk and Jenny ja Antti Wihurin Rahasto. Parts of this paper were completed while the first author was employed by the Low Temperature Laboratory in Helsinki University of Technology. Therefore this work was also supported by the Sigrid Jusélius Foundation and the Academy of Finland National Programme of Excellence 2006–2011. The authors would like to thank Jarkko Venna, Kai Puolamäki, Jukka Hyönä, Samuel Kaski, Kenneth Holmqvist and Erik D. Reichle together with anonymous reviewers for valuable comments and discussions on the manuscript.

### References

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19(6):716–723.

Bengio, Y. (1996). Input/output HMMs for sequence processing. *IEEE Trans-*

*actions on Neural Networks*, 7(5):1231–1249.

Bengio, Y. and Frasconi, P. (1999). Markovian models for sequential data. *Neural Computing Surveys*, 2:129–162.

Carpenter, R. H. S. and McDonald, S. A. (2007). Later predicts saccade latency distributions in reading. *Experimental Brain Research*, 177(2):176–183.

Carver, R. (1990). *Reading Rate: A Review of Research and Theory.* Academic Press, Inc., San Diego, California.

Cover, T. M. and Thomas, J. A. (1991). *Elements of Information Theory.* Wiley, New York.

Efron, B. and Tibshirani, R. (1993). *An Introduction to the Bootstrap.* Chapman&Hall, New York.

Feng, G. (2003). From eye movement to cognition: Toward a general framework of inference. comment on liechty et al. 2003. *Psykometrika*, 68:551–556.

Feng, G. (2006). Eye movements as time-series random variables: A stochastic model of eye movement control in reading. *Cognitive Systems Research*, 7:70–95.

Findlay, J. M. and Walker, R. (1999). A model of saccade generation based on parallel processing and competitive inhibition. *Behavioral & Brain Sciences*, 22:661–721.

Gallinari, P. (1998). Predictive models for sequence modelling, application to speech and character recognition. In Giles, C. L. and Gori, M., editors, *Adaptive Processing of Sequences and Data Structures: International Summer School on Neural Networks*, volume 1387 of *Lecture Notes in Computer Science*, pages 418–434. Springer-Verlag, Berlin, Germany.

Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B. (2003). *Bayesian Data Analysis (2nd edition).* Chapman & Hall/CRC, Boca Raton, FL.

Hastie, T., Tibshirani, R., and Friedman, J. (2001). *The Elements of Statistical Learning*. Springer, New York.

Hyönä, J., Lorch, R., and Kaakinen, J. (2002). Individual differences in reading to summarize expository text: Evidence from eye fixation patterns. *Journal of Educational Psychology*, 94:44–55.

Hyrskykari, A., Majaranta, P., Aaltonen, A., and Räihä, K.-J. (2000). Design issues of idict: a gaze-assisted translation aid. In *Proceedings of Eye Tracking Research and Applications (ETRA2000)*, pages 9–14. ACM press.

Hyrskykari, A., Majaranta, P., and Räihä, K.-J. (2003). Proactive response to eye movements. In Rauterberg, G. W. M., Menozzi, M., and Wesson, J., editors, *INTERACT'03*. IOS press.

Just, M. and Carpenter, P. (1980). A theory of reading: From eye fixations to comprehension. *Psychological review*, 87(4):329–354.

Liechty, J., Pieters, R., and Wedel, M. (2003). Global and local covert visual attention: Evidence from a Bayesian hidden Markov model. *Psychometrika*, 68:519–541.

McConkie, G. W. and Yang, S.-N. (2003). How cognition affects eye movements during reading. In J. Hyönä, R. Radach, H. D., editor, *The Mind's Eye: Cognitive and Applied Aspects of Eye Movement Research*, pages 413–427. Elsevier, Amsterdam, The Netherlands.

Miloslavsky, M. and van der Laan, M. J. (2002). Fitting of mixtures with unspecified number of components using cross validation distance estimate. *Computational Statistics and Data analysis*, 41:413–428.

Minka, T. (2000). Estimating a Dirichlet distribution. Unpublished but available in Web.

Nádas, A. (1983). A decision theoretic formulation of a training problem in speech recognition and a comparison of training by unconditional versus conditional maximum likelihood. *IEEE transactions on Acoustics, Speech, and Signal Processing*, 31(4):814–817.

O'Regan, J. K. (1990). Eye movements and reading. In Kowler, E., editor, *Eye movements and their role in visual and cognitive processes*, pages 395–453. Elsevier, Amsterdam, The Netherlands.

O'Regan, J. K. (1992). Optimal viewing position in words and the strategy-tactics theory of eye movements in reading. In Rayner, K., editor, *Eye movements and visual cognition: Scene perception and reading*, pages 333–354. Springer Verlag, New York.

Pieters, R., Rosbergen, E., and Wedel, M. (1999). Visual attention to repeated print advertising: A test of scanpath theory. *Journal of Marketing Research*, 36:424–438.

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

Pollatsek, A., Reichle, E. D., and Rayner, K. (2006). Tests of the E-Z reader model: Exploring the interface between cognition and eye-movement control. *Cognitive Psychology*, 52:1–56.

Povey, D., Woodland, P., and Gales, M. (2003). Discriminative MAP for acoustic model adaptation. In *IEEE International Conference on Acoustics, Speech, and Signal Processing, 2003. Proceedings. (ICASSP'03)*, volume 1, pages 312–315.

Puolamäki, K., Salojärvi, J., Savia, E., Simola, J., and Kaski, S. (2005). Combining eye movements and collaborative filtering for proactive information retrieval. In Marchionini, G., Moffat, A., Tait, J., Baeza-Yates, R., and Ziviani, N., editors, *SIGIR '05: Proceedings of the 28th annual international ACM SIGIR conference on Research and development in information retrieval*, pages 146–153. ACM press, New York, NY, USA.

Rabiner, L. R. (1989). A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of the IEEE*, 77(2):257–286.

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, 124(3):372–422.

Rayner, K. and Fischer, M. H. (1996). Mindless reading revisited: Eye movements during reading and scanning are different. *Perception and Psychophysics*, 58:734–747.

Rayner, K. and Pollatsek, A. (1989). *The psychology of reading*. Prentice-Hall Inc., New Jersey, USA.

Reichle, E. D., Pollatsek, A., and Rayner, K. (2006). E-Z reader: A cognitive-control, serial-attention model of eye-movement behavior during reading. *Cognitive Systems Research*, 7:4–22.

Reilly, R. G. and Radach, R. (2006). Some empirical tests of an interactive activation model of eye movement control in reading. *Cognitive Systems Research*, 7:34–55.

Richter, E., Engbert, R., and Kliegl, R. (2006). Current advances in swift. *Cognitive Systems Research*, 7:23–33.

Robertson, A. W., Kirshner, S., and Smyth, P. (2004). Downscaling of daily rainfall occurrence over northeast brazil using a hidden markov model. *Journal of Climate*, 17(22):4407–4424.

Salojärvi, J., Puolamäki, K., and Kaski, S. (2005a). Expectation maximization algorithms for conditional likelihoods. In Raedt, L. D. and Wrobel, S., editors, *Proceedings of the 22nd International Conference on Machine Learning (ICML-2005)*, pages 753–760, New York, USA. ACM press.

Salojärvi, J., Puolamäki, K., and Kaski, S. (2005b). Implicit relevance feedback from eye movements. In Duch, W., Kacprzyk, J., Oja, E., and Zadrozny, S., editors, *Artificial Neural Networks: Biological Inspirations –*

*ICANN 2005*, Lecture Notes in Computer Science 3696, pages 513–518, Berlin, Germany. Springer-Verlag.

Salojärvi, J., Puolamäki, K., and Kaski, S. (2005c). On discriminative joint density modeling. In Gama, J., Camacho, R., Brazdil, P., Jorge, A., and Torgo, L., editors, *Machine Learning: ECML 2005*, Lecture Notes in Artificial Intellligence 3720, pages 341–352, Berlin, Germany. Springer-Verlag.

Schlüter, R. and Macherey, W. (1998). Comparison of discriminative training criteria. In *Proc. ICASSP'98*, pages 493–496.

Schwartz, G. (1978). Estimating the dimension of a model. *Annals of Statistics*, 6(2):461–464.

Shimojo, S., Simion, C., Shimojo, E., and Scheier, C. (2003). Gaze bias both reflects and influences preference. *Nature Neuroscience*, 6(12):1317–1322.

Vitu, F., O'Regan, K., Inhoff, A. W., and Topolski, R. (1995). Mindless reading: Eye-movement characteristics are similar in scanning letter strings and reading texts. *Perception & Psychophysics*, 57:352–364.

Yang, S.-N. (2006). An oculomotor-based model of eye movements in reading: The competition/interaction model. *Cognitive Systems Research*, 7:56–69.

Yang, S.-N. and McConkie, G. W. (2005). New directions in theories of eye-movement control during reading. In Underwood, G., editor, *Cognitive processes in eye guidance*, pages 105–130. Oxford University Press, Great Britain.

## 6 Figure legends

**Figure 1.** An example stimulus presenting a question-answer task. The sentences are translated. The solid time line represents the time slot when the participants were instructed to find the relevant line and their eye movements were recorded. Participants proceeded in a self-paced manner, and the next trial began immediately after they typed in the line number corresponding to the selected line.

**Figure 2.** The transition probabilities and topology of the discriminative hidden Markov model. Hidden states are denoted by circles, transitions among hidden states by arrows, along with their probabilities. The beginning of the sequence is denoted by $\pi$. The capital letters on the right denote the sections of the HMM that were assigned for each of the tasks (W=word search, A=question-answer, I=true interest), small letters within the hidden states denote the names of the hidden states, (s=*scanning*,r=*reading*,d=*decision*).

**Figure 3.** Examples of eye movement trajectories in the experiment. The HMM states along the most probable paths are denoted by 'x' – state 1 (*scanning*), '△' – state 2 (*reading*), □ – state 3 (*decision*).See text for interpretations of the states. The beginning of the trajectory is marked with a circle; ending with two concentric circles. W: word search. A: question – answer. I: True interest.

**Figure 4.** Average probability (y-axis) of being in state $s$. Horizontal axis is the normalized sequence length. Top row: word search (W). Center: question-answer (A). Bottom: true interest (I). The plots show the mean probability (and ± one standard deviation; 66 % confidence interval) of being in a given HMM state as a function of time. Left column: state 1, middle column: state 2, right column: state 3.

Table 1

Confusion matrix from the test data, showing the number of assignments classified by the logistic regression into the three task types (columns) versus their true task type (rows). The diagonal contains the number of correctly predicted assignments. The percentages (in parentheses) denote row- and column-wise classification accuracies. The row-wise accuracy shows the percentage of correctly predicted assignments for the given task type, the column-wise accuracy shows the percentage of correctly predicted task types, given the prediction.

|  | Prediction |  |  |
|---|---|---|---|
|  | $W$ (66.2 %) | $A$ (45.3 %) | $I$ (60.0 %) |
| $W$ (77.2 %) | 139 | 23 | 18 |
| $A$ (28.3 %) | 55 | 43 | 54 |
| $I$ (70.6 %) | 16 | 29 | 108 |

Table 2

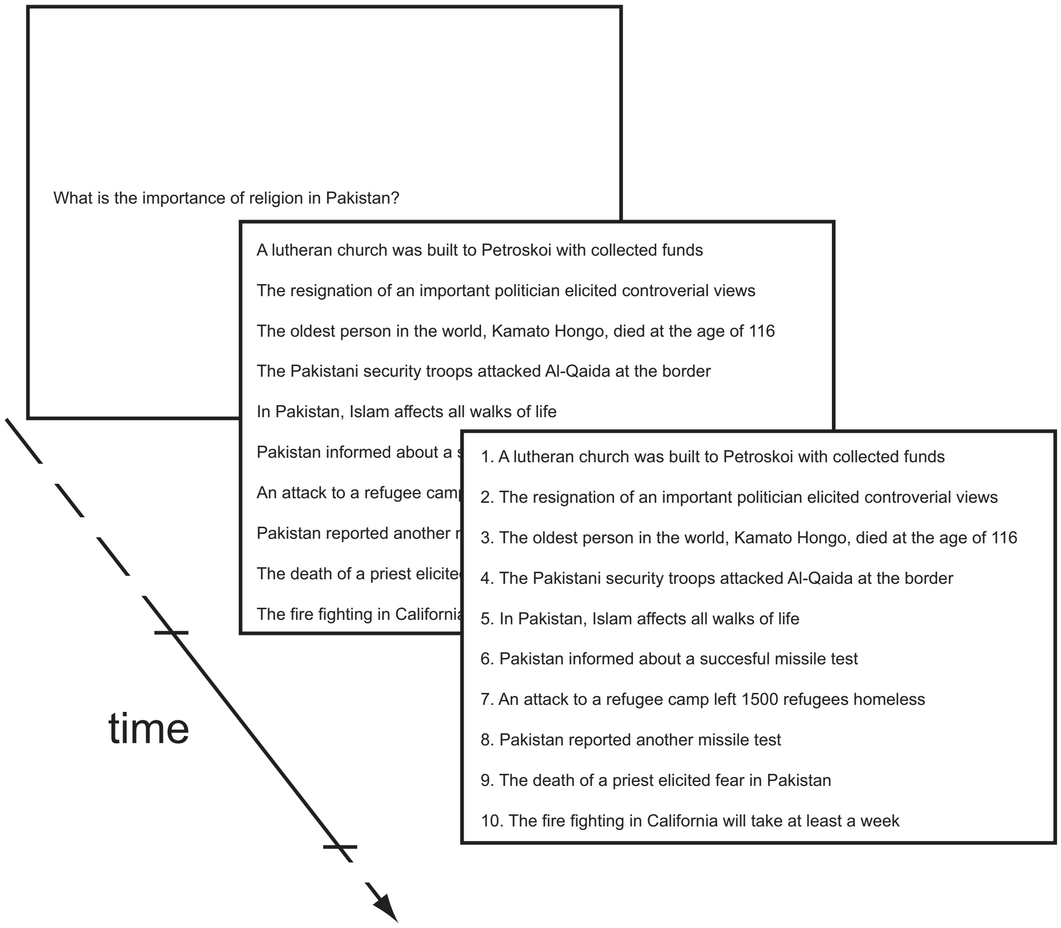
Confusion matrix showing the number of assignments classified by the discriminative HMM into the three task types (columns) versus their true task type (rows). The percentages (in parentheses) denote row- and column-wise classification accuracies.

| | | Prediction | |
|---|---|---|---|
| | $W$ (70.0 %) | $A$ (50.0 %) | $I$ (57.5 %) |
| $W$ (78.9 %) | 142 | 22 | 16 |
| $A$ (35.5 %) | 43 | 54 | 55 |
| $I$ (62.8 %) | 18 | 39 | 96 |

Table 3

Discriminative HMM parameter values for scanning, reading and decision states for each task type (corresponding maximum likelihood estimates in parentheses). In saccade lengths, 160 pixels approximates to 13 letters. Standard deviation $\sigma$ is reported with respect to mean $\mu$ by $\left[{\mu-\sigma \atop \mu+\sigma}\right.$, where applicable (67 % of the probability mass is within this interval).

| | | **Scanning** | **Reading** | **Decision** |
|---|---|---|---|---|
| Probability of | Word search | 32 % (17) | 16 % (15) | 0 % |
| beginning the task | Question – answer | 20 % (21) | 10 % (12) | 0 % |
| | True Interest | 15 % (17) | 7 % (8) | 0 % |
| **Word Search** | Fixation duration (ms) | $134\left[{100 \atop 180}\right.$ (125) | $199\left[{140 \atop 284}\right.$ (187) | $171\left[{92 \atop 320}\right.$ (219) |
| –observations | Saccade length (pix) | $166\left[{68 \atop 409}\right.$ (155) | $132\left[{67 \atop 259}\right.$ (159) | $132\left[{54 \atop 319}\right.$ (120) |
| | Saccade direction: forward | 31 % (34) | 61 % (53) | 39 % (22) |
| | upward | 22 % (21) | 6 % (9) | 6 % (3) |
| | backward | 19 % (16) | 15 % (15) | 20 % (36) |
| | downward | 28 % (28) | 18 % (23) | 17 % (2) |
| | End assignment | 1 % (0) | 0 % | 18 % (37) |
| | Previous fixations=true | 23 % (25) | 24 % (15) | 78 % (64) |
| **Question – answer** | Fixation duration (ms) | $134\left[{99 \atop 182}\right.$ (129) | $205\left[{141 \atop 299}\right.$ (204) | $177\left[{96 \atop 323}\right.$ (173) |
| –observations | Saccade length (pix) | $160\left[{60 \atop 422}\right.$ (156) | $133\left[{74 \atop 239}\right.$ (141) | $137\left[{48 \atop 391}\right.$ (133) |
| | Saccade direction: forward | 37 % (39) | 63 % (63) | 33 % (35) |
| | upward | 21 % (20) | 5 % (5) | 14 % (15) |
| | backward | 16 % (14) | 12 % (12) | 27 % (26) |
| | downward | 27 % (26) | 20 % (20) | 10 % (12) |
| | End assignment | 0 % | 0 % | 16 % (11) |
| | Previous fixations=true | 28 % (25) | 26 % (21) | 86 % (83) |
| **True Interest** | Fixation duration (ms) | $134\left[{97 \atop 184}\right.$ (125) | $200\left[{138 \atop 291}\right.$ (196) | $176\left[{95 \atop 326}\right.$ (169) |
| –observations | Saccade length (pix) | $160\left[{57 \atop 452}\right.$ (165) | $128\left[{73 \atop 226}\right.$ (131) | $133\left[{48 \atop 365}\right.$ (135) |
| | Saccade direction: forward | 41 % (43) | 61 % (61) | 37 % (38) |
| | upward | 21 % (19) | 7 % (7) | 15 % (16) |
| | backward | 13 % (11) | 13 % (14) | 30 % (25) |
| | downward | 26 % (26) | 19 % (18) | 11 % (14) |
| | End assignment | 0 % | 0 % | 7 % (8) |
| | Previous fixations=true | 27 % (28) | 24 % (26) | 86 % (88) |



What is the importance of religion in Pakistan?

A lutheran church was built to Petroskoi with collected funds

The resignation of an important politician elicited controverial views

The oldest person in the world, Kamato Hongo, died at the age of 116

The Pakistani security troops attacked Al-Qaida at the border

In Pakistan, Islam affects all walks of life

Pakistan informed about a s

An attack to a refugee camp

Pakistan reported another m

The death of a priest elicited

The fire fighting in California

1. A lutheran church was built to Petroskoi with collected funds

2. The resignation of an important politician elicited controverial views

3. The oldest person in the world, Kamato Hongo, died at the age of 116

4. The Pakistani security troops attacked Al-Qaida at the border

5. In Pakistan, Islam affects all walks of life

6. Pakistan informed about a succesful missile test

7. An attack to a refugee camp left 1500 refugees homeless

8. Pakistan reported another missile test

9. The death of a priest elicited fear in Pakistan

10. The fire fighting in California will take at least a week

Fig. 1.

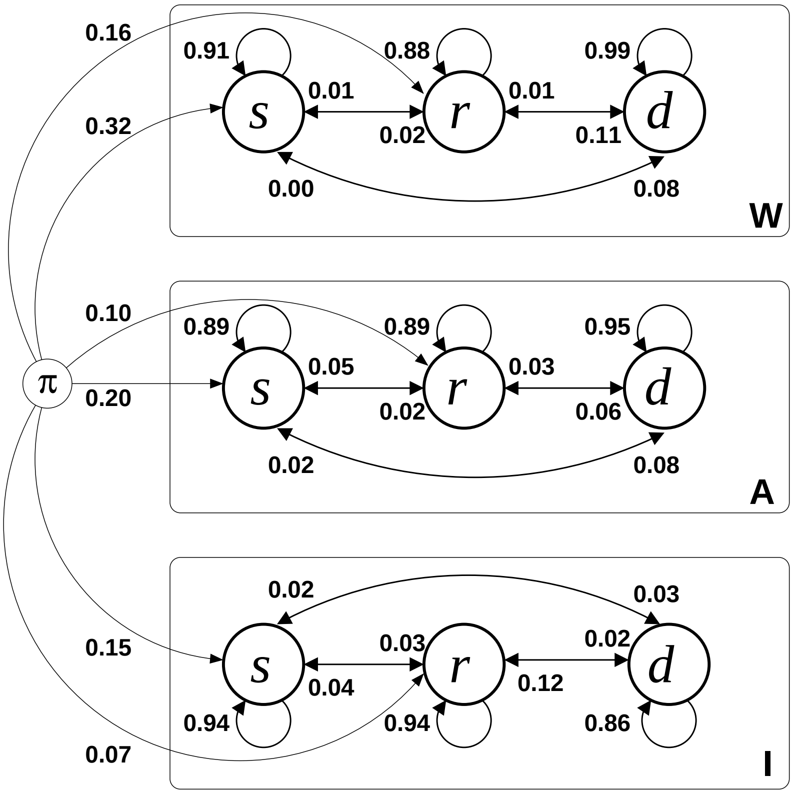

0.16
0.91
0.88
0.99
0.01
0.01
0.32
s
r
d
0.02
0.11
0.00
0.08
W
0.10
0.89
0.89
0.95
π
0.05
0.03
0.20
s
r
d
0.02
0.06
0.02
0.08
A
0.02
0.03
0.03
0.02
0.15
s
r
d
0.04
0.12
0.94
0.94
0.86
0.07
I

Fig. 2.

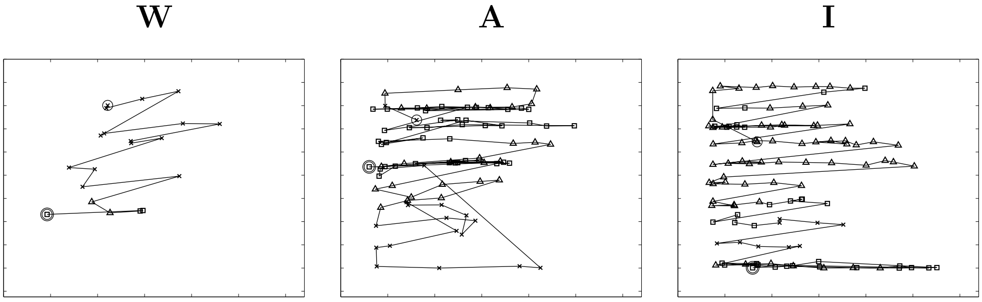

W
A
I

Fig. 3.

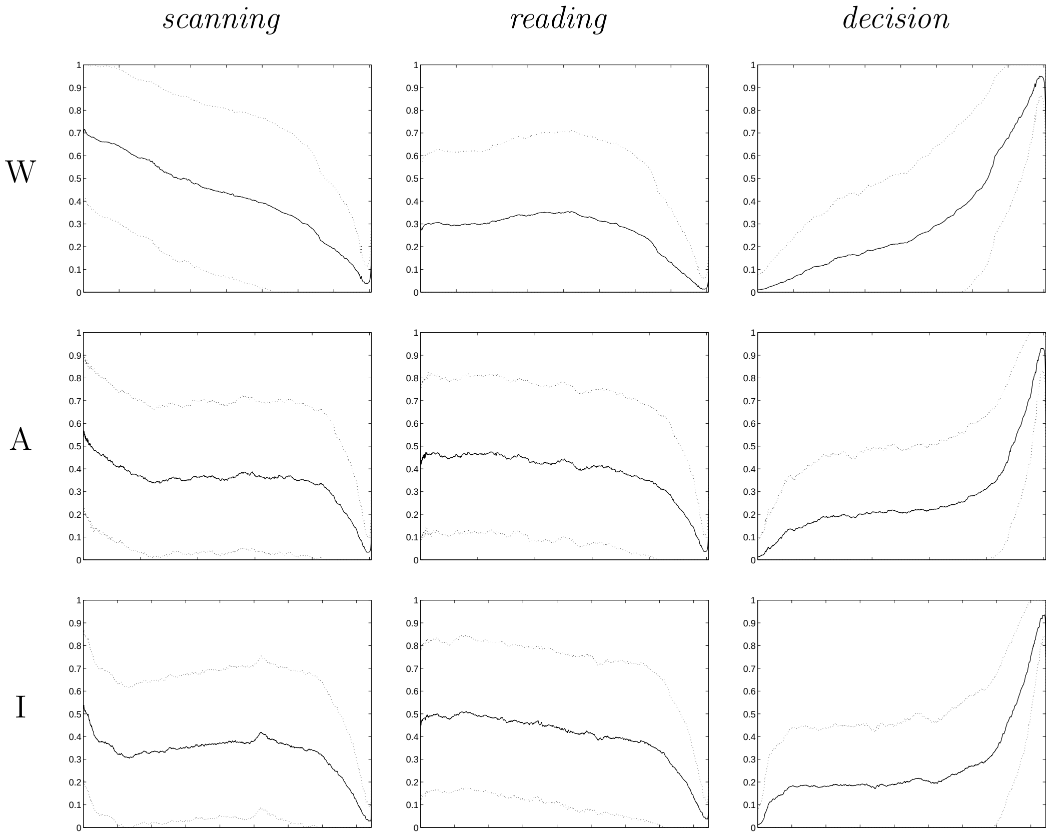


Fig. 4.

Table 4

Expected dwell times and standard deviations in scanning, reading and decision states, plus times before and after reaching the decision state, along with the mean percentages of prevalence of the states. Values are computed from the observation trajectory which was segmented using the Viterbi algorithm on dHMM. Capital letters denote the tasks (W = word search, A = question-answer, I = true interest), and units are in seconds. Error estimates (±) are 95 % confidence intervals, obtained with a bootstrap method with 400 replicate data sets.

| | **W** | | **A** | | **I** | |
|---|---|---|---|---|---|---|
| | mean | stdev. | mean | stdev. | mean | stdev. |
| Total $T$ | 4.1±0.4 | 3.1±0.5 | 8.5±1.2 | 7.1±1.8 | 11.6±1.1 | 6.7±0.9 |
| $T$ in *scanning* | 2.2±0.3 | 1.8±0.4 | 2.8±0.4 | 2.3±0.3 | 3.4±0.4 | 2.4±0.2 |
| $T$ in *reading* | 4.3±0.7 | 3.2±0.7 | 6.1±1.0 | 5.1±0.8 | 6.2±0.8 | 5.2±0.6 |
| $T$ in *decision* | 0.7±0.1 | 0.8±0.4 | 1.4±0.4 | 2.9±1.8 | 1.8±0.3 | 2.0±0.4 |
| $T$ to *decision* | 3.4±0.4 | 2.8±0.5 | 6.1±0.7 | 4.5±0.6 | 8.0±0.7 | 4.6±0.7 |
| $T$ after *decision* | 0.8±0.2 | 1.0±0.4 | 2.5±0.7 | 4.8±1.6 | 3.6±0.8 | 5.1±1.0 |
| % in *scanning* | 51±6 | 40±2 | 47±6 | 40±2 | 47±6 | 40±2 |
| % in *reading* | 33±6 | 41±2 | 38±6 | 41±2 | 38±6 | 41±2 |
| % in *decision* | 16±2 | 15±2 | 15±2 | 15±2 | 15±2 | 15±2 |