Estimation of Browsing States in Consumer Decision Processes from Eye Movements

Erina Ishikawa Schaffer, Hiroaki Kawashima, and Takashi Matsuyama
Kyoto University
Yoshida-Honmachi, Sakyo-ku Kyoto 606-8501, Japan
ishikawa@vision.kuee.kyoto-u.ac.jp

Abstract

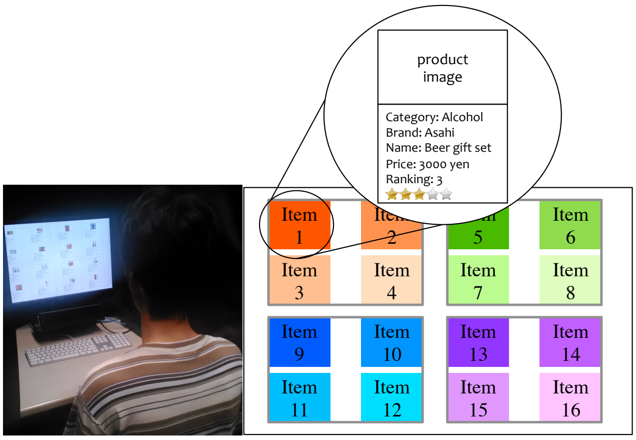
Eye movements can be an important cue to reveal consumer decision processes. Findings from existing studies suggest that the consumer decision process consists of a few different browsing states such as screening and evaluation. To reveal the characteristics and temporal changes of browsing states in catalog browsing situations, this study proposes a hidden semi-Markov-based gaze model, where the states in consumer decision processes are explicitly modeled as hidden states. To achieve a better understanding of consumer decision processes, eye movements are first encoded to a sequence of gaze features using semantic structure of digital catalogs. The proposed gaze model is trained and evaluated using gaze data which was collected from eight participants in an experimentally controlled catalog browsing situation. We analyze the estimated browsing states and demonstrate that the states can contribute to improve the performance of viewer interest estimation.

1. Introduction

Observing the behavior of consumers is fundamental to understanding their decision processes. Eye movements have been considered a good indicator of the processing states of the decision process since they directly represent how humans process information about displayed items [8, 3]. From the observation of eye movements, Russo et al. insist that choice processes start from an *orientation* stage, followed by an *evaluation* stage, and finally a *verification* stage [8]. They define the *orientation* stage to be the stage that occurs before the first comparison pattern (the re-fixation pattern of the form X-Y-X...) appears, the *evaluation* stage is to be between the first and the last comparison pattern, and the *verification* stage to be after the last comparison pattern. However, their three-stage model cannot always be applied to real-world environments since consumer decision processes can be affected by their individual goals or tasks [3]. Moreover, the transitions among stages are not always one-way; for example, it is more likely for consumers to go back to the *orientation* stage to re-input item information after the *evaluation* stage.

Unlike previous approaches, we model consumers' decision process with probabilistic transitions among the processing states and estimate the processing states from eye movements in a bottom-up manner. To deal with complex eye-mind link, probabilistic approaches are often introduced in previous gaze models [11, 6]. Simola et al. [10] take a similar approach to ours: they apply a hidden Markov model (HMM) to sequences of motion features of eye movements (e.g., saccade lengths) to uncover processing states of information-search in sentence reading. Three HMMs were built corresponding to three different information-search tasks: simple word search, finding the answer to a question, and selecting an interesting title from a list. Their results showed that the tasks can be estimated from a newly observed sequence of eye movements by comparing the likelihood of each model.

Although Simola's model is effective for analyzing information-search behavior, it cannot be directly applied to consumer decision processes. We are particularly interested in the interpretation of gaze behavior that corresponds to the processing states of decision processes, which we call *browsing states*. In this paper, we assume the situation where a viewer is browsing a digital catalog on a computer display, and we address the problem of building a gaze model whose states can be associated with the findings from existing studies. We make two key observations: semantic information of digital catalogs influences viewers' eye movements, and the amount of time a viewer spends in a state is an important feature of that state [8]. Therefore, we incorporate (1) information about catalog content and (2) the duration of browsing states into our model. To incorporate (1), we use the *designed structure* of digital catalogs to encode eye movements into a sequence of gaze features. Designed structure, as proposed in [4], is high-level content structure that reflects the designer's point of view,



(a) Apparatuses. (b) Layout of digital catalogs

Figure 1. Experimental setting.

such as groupings of items. Instead of using semantic attributes of items, the designed structure employs embedded semantic information as a content layout. To incorporate (2), this study introduces the use of a hidden semi-Markov model (HSMM, [12]) to represent the relationships between browsing states and encoded gaze features. HSMM is an extended model of HMM that each state has a duration distribution explicitly, which is often used to model human behavior in various context [1]. Understanding viewers' browsing states from their behavior has rich potential for various applications, such as interactive information systems that provide information with proper timing.

To evaluate how the proposed model estimates meaningful hidden browsing states behind the gaze sequences, we collected gaze data in a catalog browsing situation with a procedure explained in Sec. 2. The data is later used for the training of the model (Sec. 3) and its evaluation (Sec. 4).

2. Collection of gaze data

Eight participants took part in the experiment. Each participant was asked to sit in front of a screen showing a digital catalog (see Fig. 1 (a)). Gaze data of the participants were acquired as 2D points on the screen by using an eye tracker [1] installed below the screen.

2.1. Digital catalogs

For each participant, eight digital catalogs were prepared. Each digital catalog contained the description (images and text) of 16 items (see Fig. 1 (b)). The semantic attributes and attribute values of items that were considered in this study are listed in Table 1. The semantic information was explicitly described in text on the catalogs so that viewers could understand it without prior knowledge. The items in each catalog were grouped by either *price* or *category* attribute. The item positions within a group were randomized every time the catalog was shown to a participant.

Table 1. The attributes and attribute values used in the experiments.

Category	Price (yen)	Ranking	Review
Delicatessen	1001-3000	1-4th	1-star
Sweets	3001-5000	11-14th	2-star
Alcohol	5001-7000	21-24th	3-star
Household goods	7001-	31-34th	4-star
			5-star

2.2. Procedure

Participants were asked to select an item from a catalog as a seasonal gift. The situation of selection can vary based on participants' preference and products in catalogs. For example, it is possible that participants make a decision among multiple candidates, or perhaps they can find only one product that satisfies their criteria. Since it is difficult to obtain such information about the situations, we gave participants a request that specified the requirements for items to select for each trial, for example: "Select an item in *Alcohol* category and with more than 4-star review." The request was controlled for the number of items that the participants considered as alternatives. As noted in [7], goal-directed search behavior can be induced in participants during in-store experiments by fixing requests so that only one item in the store satisfies all of the conditions. We take a similar approach here, however, we change the number of items that satisfy the requests to be between 0 to 3 to induce browsing behavior where a decision must be made among multiple alternatives.

The participants were instructed to follow the request to the best of their ability. During catalog browsing, participants were able to refer to the request by pressing the space key on a keyboard that was provided. The gaze data during catalog browsing were registered together with the time stamps when the space key was pressed. After deciding an item, participants were asked to submit any items they considered as candidates alongside their final selection.

2.3. Examples of gaze data

Eye trackers sometimes contain noise or miss viewers' gaze points. Therefore, the sequence of gaze points was first smoothed by applying a median filter [2]. For the analysis, a sequence of gaze points were first converted to a sequence of item regions being looked at. In gaze region sequences, if successive intervals with the same item ID were interrupted by a blink, the intervals were combined to a longer interval. The gaze region sequences were also modified by dis-

[1]Tobii X120 (freedom of head movement: 400x220x300 mm, sampling rate: 60 Hz, accuracy: 0.5 degrees)

[2]In this paper, the window size of the median filter was 5 sampling points at 60 Hz (corresponding to about 83 msec).

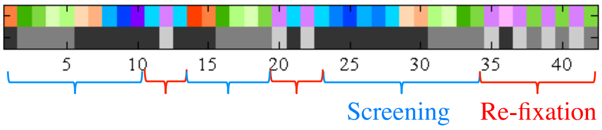


Figure 2. Example of gaze region sequences (the 4th trial of the 8th participant). In the top row, the color corresponds to the color of regions in Fig. 1 (b). In the bottom row, the selected item and candidates are shown by light gray and gray highlights respectively (corresponding to the color used in Fig. 7).

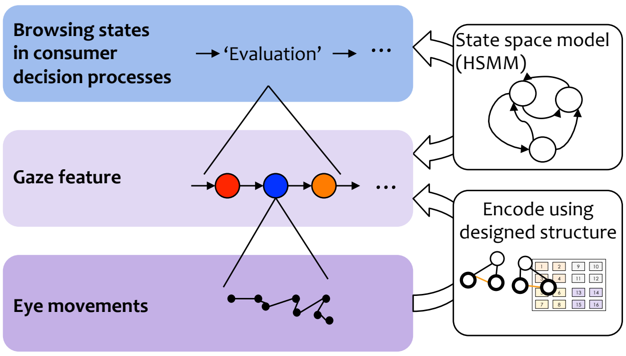


Figure 3. Overview of the proposed method.

carding intervals shorter than a threshold (100 msec). Any interval in which the participant was referring to the requirements of the request was also discarded.

An example of gaze region sequences is shown in Fig. 2. Re-fixation patterns of the form X-Y-X... and screening patterns of the form X-Y-Z... appear alternatively in the example. This example highlights the importance of considering probabilistic transitions between browsing states to represent catalog browsing behavior.

3. Hidden semi-Markov-based gaze model

The overview of the proposed model is shown in Fig. 3. The eye movements are first encoded using the designed structure of digital catalogs (Sec. 3.1), then a HSMM is applied to the sequence of encoded gaze features (Sec. 3.2).

3.1. Sequence of gaze feature

3.1.1 Modeling the content structure

We employ *designed structure* proposed in [4] to encode eye movements into gaze features. The designed structure can be represented as follows. Suppose a catalog contains information of a set of items $\mathcal{I}_{All} = \{1, \dots, K\}$. Each item has a set of P attributes $\mathcal{P}_{All} = \{1, \dots, P\}$, where p-th attribute can take a value of A_p possible attribute values $\mathcal{A}^{(p)} = \{1, \dots, A_p\}$. Here, let us introduce a function $f_p : \mathcal{I}_{All} \to \mathcal{A}^{(p)}$, where $f_p(i)$ indicates the attribute value of the p-th attribute that the i-th item has. In the content layout used in the experiment, the designer aims to emphasize a specific attribute (e.g., *category*), then all items in the same category are regarded as "in the same group" and allocated based on their category type. To represent this process, a set of the emphasized attributes is described as $\mathcal{P}_E \subset \mathcal{P}_{All}$. Using $\mathcal{P}_E$, the relations among items in the digital catalogs can be determined as follows.

Parallel relation. Two different items i and j have this relation when the items share one or more emphasized attributes; that is, $f_p(i) = f_p(j), \exists p \in \mathcal{P}_E$.

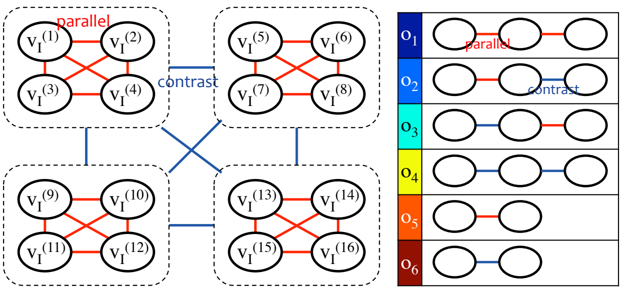
Contrast relation. Two different items i and j have this relation when the items do not share any emphasized attributes; that is, $f_p(i) \neq f_p(j), \forall p \in \mathcal{P}_E$.

In [4], the versatility of the use of designed structure is confirmed by using multiple types of layouts. In this paper, we focus on analysis of the sequence of encoded gaze features. Therefore, we only use catalogs with category-based layout that can be simply described by the above two relations. We represent the relations among items as a directed graph for the encoding of eye movements (see Sec. 3.1.2). Let us denote a set of nodes $V_I = \{v_I^{(k)} \mid k = 1, \dots, K\}$ corresponding to the K items. Here, design-relation edges are defined as $E_D \subseteq V_I \times V_I$, where each edge $e_d \in E_D$ has either label *parallel* or *contrast*. Finally, the designed structure is defined by a directed graph $G_D = (V_I, E_D)$ (as shown in Fig. 4 (a)).

3.1.2 Encoding eye movements into gaze features

Gaze data are obtained as sequences of gaze points on the screen. Let us denote item regions in a digital catalog as $\mathcal{R} = \{R_1, \dots, R_K\}$. A sequence of gaze points is first converted to a sequence of item regions being looked at $r = (r_0, \dots, r_J)$ $(r_j \neq r_{j+1}, r_j \in \mathcal{R})$.

The gaze region sequence, $r = (r_0, \dots, r_J)$, is associated with design relation labels derived from the graph of



(a) Designed structure of digital catalogs used in the experiment.

(b) Gaze features.

Figure 4. Encoding of eye movements. (a) Edges between dotted frames indicate that every node in the frame is connected to every node in the other frame. (b) Six patterns of tri-gram paths in the graph G_D. The two bottom features correspond to re-fixations.

the designed structure, G_D. For each tri-gram of gaze regions $(r_{j-1}, r_j, r_{j+1})(j \in [1, J-1])$ in a gaze region sequence, a path of corresponding item nodes in the graph G_D is obtained. By considering only the labels of edges of each path, we can categorize the paths into six patterns as shown in Fig. 4 (b). Note that the four top features in Fig. 4 (b) correspond to tri-grams with three different gaze regions, that is, $r_{j-1} \neq r_{j+1}$, and the two bottom ones correspond to re-fixations, that is, $r_{j-1} = r_{j+1}$.

3.2. Learning the model parameters

HSMM (also known as explicit duration hidden Markov model) is an extension of HMM that considers an additional parameter for explicitly modeling the duration of states. Let us denote the original HMM as $\lambda = (Q, O, A, B, \pi)$, where O is a set of possible observations $O = \{o_1, \ldots, o_N\}$, Q is a set of hidden states $Q = \{q_1, \ldots, q_M\}$, A is the $M \times M$ state transition probability matrix, B is $M \times N$ emission probability matrix, and π is the initial probability of the states. Note that, in this study, the set O corresponds to the set of gaze feature labels (Fig. 4 (b)).

HSMM can be represented as $\lambda = (Q, O, A, B, P, \pi)$ by adding the $M \times D$ duration distribution matrix P to the original HMM, where D is the maximum state duration and $p_{m,d} \in C$ indicates the probability that the m-th state last for d-time units. For learning the model parameters, this study uses the efficient EM algorithm proposed in [12]. Since the data is limited, the duration distribution, P, is approximated by a Gaussian distribution.

In the EM iteration, for the initial values of A and π in HMM and A, P, and π in HSMM, uniform distributions are used. For the initial values of B in both HMM and HSMM, distributions of the gaze features are first obtained by applying a sliding window of size 20. Then, each mean point of M clusters of the distributions obtained by k-means clustering are used.

To decide the number of states M for HSMM and HMM, we utilized Akaike Information Criterion (AIC) since the sample size is limited against the model complexity [2]. AIC can be represented as $AIC = -2\log L(\hat{\lambda}) + 2h$, where $\log L(\hat{\lambda})$ is the likelihood of the model with the estimated parameters $\hat{\lambda}$, and h is the degree of freedom of model parameters. The best number of states can be decided by finding M that minimizes the value of AIC.

4. Evaluation

Ideally, the ground truth about the viewers' browsing state could be used to evaluate the validity of the proposed model. However, since such ground-truth data are not available, we instead use an estimation of the viewers' interest to evaluate the trained models. First, we discuss the interpretation of estimated states using the collected gaze data in Sec. 2. Then, we show how the states can be used for viewer's interest estimation.

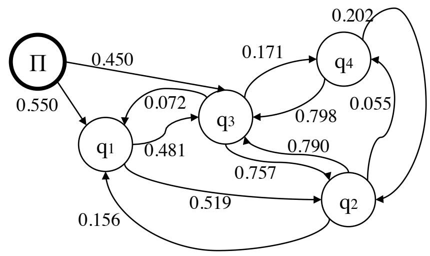


Figure 5. The topology of the trained HSMM.

4.1. Interpretation of estimated browsing states

Using all the collected gaze data, we trained an HSMM, where $M = 4$ was selected with AIC (Sec.3.2). The topology of the trained HSMM, and emission probabilities are shown in Fig. 5 and Fig. 6, respectively. Fig. 5 suggests that a viewer tends to start in state q_1 or q_3, then shift to state q_2 and q_4. Fig. 7 shows which kind of items are more likely to be looked at in each estimated state. The items are categorized into selected items, candidates, and others (non-selected).

From Fig. 6, state q_1 or q_3 can be interpreted as a state in which a viewer is more likely to look at items in the same group. From Fig. 7, viewers look at fewer selected items/candidates in these states. Meanwhile, in state q_2 and q_4, viewers are more likely to shift their gaze actively among items in different groups and look at selected items/candidates. To associate the obtained states with the three-stage model [8], we can consider state q_1 or q_3 to be *orientation* or *verification* (these two stages share a similar pattern, but can be discriminated based on context), while state q_2 and q_4 can be interpreted as *evaluation*.

The posterior probabilities of browsing states can be calculated with a newly observed gaze sequence and the trained HSMM. An example timeline of browsing states is shown in Fig. 8. The average durations of states in HSMM were q_1: 11.75, q_2: 4.39, q_3: 9.18, and q_4: 4.80, which are longer than the average durations of states in HMM (q_1: 1.81, q_2: 2.72, q_3: 1.00, q_4: 1.00, q_5: 1.77). Note that the lengths correspond to the number of gaze features, that is, the number of item regions looked at (refer to Sec. 3.1.2). The results show that the duration distributions of hidden states of an HSMM contribute to estimate reasonable browsing states since the processing states in consumer decision processes do not change frequently [8].

4.2. Estimation of viewers' interest

According to the findings from gaze analysis on choice behavior, visual attention to competing items increases significantly when a consumer is close to making a deci-

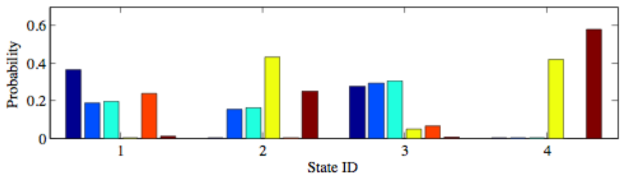


Figure 6. Emission probabilities of each state in HSMM. The color of the bars corresponds to the color of gaze features in Fig. 4 (b).

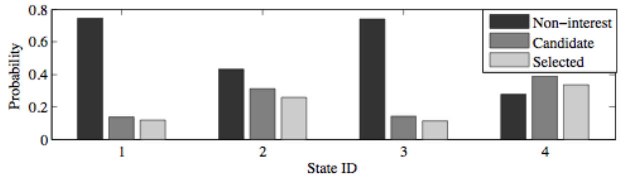


Figure 7. The probability of selected items, candidates, and others being looked at in each browsing state.

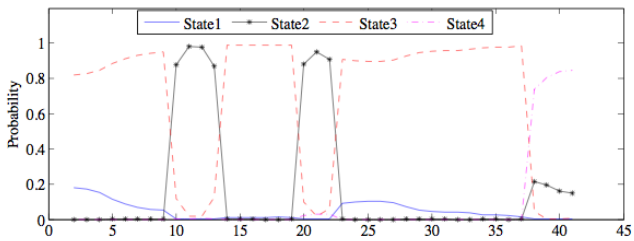


Figure 8. An example timeline of browsing states estimated by the trained HSMM for the gaze region sequence of the 4th trial of the 8th participants (refer to Fig. 2).

sion [9]. Therefore, we hypothesize that the estimation of viewer's interest can be improved by identifying when the viewer enters the *evaluation* state, and we propose an estimation method of viewers' interest from eye movements, simply by finding items that the viewer looked at for the longest duration during *evaluation* states (q_2 and q_4).

We evaluated our proposed model against three methods: HMM, Sliding window, and Baseline. Leave-one-subject-out cross validation was used; that is, gaze data of one participant was used as test data and the remaining of the data was used to train the model. By comparing the results of HSMM with HMM, we can evaluate the effectiveness of the duration distribution of browsing states. As for the number of states, $M = 5$ was the best for HMM. In Sliding window, a sliding window of size 20 is first applied to gaze feature sequences. The browsing state at j-th gaze region was determined by finding the nearest point in the distributions used as initial values of B for training HSMM to the distribution of gaze features of interval $[r_{j-9}, r_{j+10}]$. As with HSMM, *evaluation* states estimated by HMM and Sliding window were identified based on which kind of items were more likely to be looked at in each estimated state. In the

Table 2. The estimation of the selected item.

HSMM	HMM	Sliding window	Baseline
0.4844	0.4688	0.3281	0.4375

Table 3. The estimation of the set of considered items (selected items and candidates).

	Precision	Recall	F1 score
HSMM	0.3315	**0.8264**	0.4732
HMM	**0.3651**	0.7986	**0.5011**
Sliding window	0.1876	0.6944	0.2954

Baseline method, the entire gaze sequence was considered when finding the item with the longest duration, and was taken to be the selected item. Tab. 2 shows that the use of browsing states estimated by our proposed model enables the estimation of selected items with the highest accuracy.

In addition to the previous analysis, we also conducted an estimation of the set of considered items (selected items and candidates). This is because gaze behavior on selected items and candidates can be very similar in a situation where there are multiple attractive candidates in the catalog. For the estimation of considered items, we find all items looked at during *evaluation* states. Thus, the estimation results of the items of interest using Baseline method is omitted. Tab. 3 shows that the recall is the highest with our proposed model, meanwhile, the precision and F1 score is the highest with HMM.

4.3. Discussion

The results from the interpretation of estimated browsing states show that the proposed model can automatically estimate browsing states that can be interpreted as the processing states defined in existing studies.

The results from the estimation of the selected items show that considering both duration distributions and probabilistic transitions among browsing states are effective to represent gaze behavior in consumer decision processes. The results from the estimation of considered items indicate that HSMM and HMM have advantages and disadvantages for estimating considered items. This is because, *evaluation* states in HSMM have wider intervals due to the state duration distributions, therefore, the possibility of non-considered items being in the *evaluation* state can be higher. However, HSMM can detect considered items even if they do not directly relate to re-fixation patterns.

5. Conclusion

In this paper, a gaze model based on hidden semi-Markov model is proposed to understand browsing states in consumer decision processes. The proposed model enables estimation of browsing states in a bottom-up manner. We confirm the validity of estimated states through estimation of viewers' interest based on estimated states. For future work, we are considering to investigate further detailed and various viewers' states, such as consumers' strategies [5].

Acknowledgments

This work is supported by Grant-in-Aid for Scientific Research under the contract of 25·5396 and JSPS KAKENHI Grant Number 26280075.

References

[1] M. Baratchi, N. Meratnia, P. J. Havinga, A. K. Skidmore, and B. A. Toxopeus. A Hierarchical Hidden Semi-Markov Model for Modeling Mobility Data. In *The 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, 2014. 2

[2] F. Cartella, J. Lemeire, L. Dimiccoli, and H. Sahli. Hidden Semi-Markov Models for Predictive Maintenance. *Mathematical Problems in Engineering*, 2015, 2015. 4

[3] K. Gidlöf, A. Wallin, R. Dewhurst, and K. Holmqvist. Using Eye Tracking to Trace a Cognitive Process : Gaze Behaviour During Decision Making in a Natural Environment. *Eye Movement Research*, 6(1):1–14, 2013. 1

[4] E. Ishikawa, H. Kawashima, and T. Matsuyama. Using Designed Structure of Visual Content to Understand Content-Browsing Behavior. *IEICE Transactions*, 98(8), 2015. 1, 3

[5] W. W. Moe. Buying, Searching, or Browsing: Differentiating Between Online Shoppers Using In-Store Navigational Clickstream. *Consumer Psychology*, 13(1&2):29–39, 2003. 6

[6] Y. Nakano and R. Ishii. Estimating User's Engagement from Eye-gaze Behaviors in Human-Agent Conversations. In *The 15th International Conference on Intelligent User Interfaces*, pages 139–148, 2010. 1

[7] J. Pfeiffer, M. Meißner, J. Prosiegel, and T. Pfeiffer. Classification of Goal-Directed Search and Exploratory Search Using Mobile Eye- Tracking. In *The 5th International Conference on Information Systems*, pages 1–14, 2014. 2

[8] J. Russo and F. Leclerc. An Eye-Fixation Analysis of Choice Processes for Consumer Nondurables. *Consumer Research*, 21(2):274–290, 1994. 1, 4

[9] S. Shimojo, C. Simion, E. Shimojo, and C. Scheier. Gaze Bias Both Reflects and Influences Preference. *Nature Neuroscience*, 6(12):1317–1322, Dec. 2003. 5

[10] J. Simola, J. Salojärvi, and I. Kojo. Using hidden Markov Model to Uncover Processing States from Eye Movements in Information Search Tasks. *Cognitive Systems Research*, 9(4):237–251, Oct. 2008. 1

[11] R. Yonetani, H. Kawashima, and T. Matsuyama. Multi-mode Saliency Dynamics Model for Analyzing Gaze and Attention. In *Symposium on Eye Tracking Research and Applications*, volume 1, pages 115–122, 2012. 1

[12] S.-Z. Yu and H. Kobayashi. Practical Implementation of an Efficient Forward-Backward Algorithm for an Explicit-Duration Hidden Markov Model. *Signal Processing*, 54(5):1947–1951, 2006. 2, 4