Improving Gaze-based Selection using Variable Dwell Time

Zhaokang Chen and Bertram E. Shi, Fellow, IEEE

Abstract—In order to avoid the “Midas Touch” problem, gaze-based interfaces for selection often introduce a dwell time: a fixed amount of time the user must fixate upon an object before it is selected. Past interfaces have used a uniform dwell time across all objects. Here, we propose an algorithm for adjusting the dwell times of different objects based on the inferred probability that the user wishes to select them. In particular, we introduce a probabilistic model of gaze behavior and train it on natural gaze trajectories collected while users surf the web through a gaze-based web browser. Using this model, we can infer the probability that each hyperlink on the web page is the one the user intends to select. We assign shorter dwell times to more likely hyperlinks and longer dwell times to less likely hyperlinks. The resulting variable dwell time gaze-based browser enables subjects to surf the web faster and/or more reliably. We use a two-stage hidden Markov model based algorithm to model the natural eye gaze trajectories. We have evaluated this method objectively both in simulation and experimentally, and subjectively through questionnaires. Our results demonstrate that the proposed algorithm achieves a better tradeoff between accuracy and speed. This work demonstrates that natural eye movements can be used to infer user intent, and that this information can be used to improve gaze-based selection without increased cognitive load on users.

Index Terms—Human computer interface, Gaze tracking, Hidden Markov models, Inference algorithms, Browsers, Bayes methods

I. INTRODUCTION

The use and capabilities of eye trackers have expanded rapidly in recent years. A state-of-the-art eye tracker can estimate eye gaze on a monitor screen accurately enough to allow users to interact with a computer system. This is especially useful for people with motor disabilities [1]. Eye movement can be regarded as a very fast interaction modality and can be very informative about users’ intent. As a result, use of eye movements is getting more and more attention in the field of human computer interaction (HCI) [2].

Many innovative human computer interfaces have been designed using eye tracking devices. One of the most common uses of eye gaze is to select the object of interest to the user or as a pointing modality [3, 4]. However, one of the most challenging problems for this kind of interfaces is “Midas Touch” problem: it is difficult to distinguish between spontaneous eye movements for gathering visual information and intentional eye movements for explicit selection [5]. The most common way to avoid unintentional selection is to introduce a dwell time. Users must maintain their eye gaze on an object for a predefined duration before it is selected [4, 6, 7].

Many state-of-the-art gaze-based user interfaces are constructed around selection using dwell time, e.g. Windows Control [8], developed by Tobii dynavox, and GazeTheWeb [9], developed by Multimedia Authoring & Management using your Eyes & Mind (MAMEM). Both use a two-step control policy: the users first select a command by dwelling on a button at the side of the screen, and then select the object of the command by dwelling on that. Potentially, these systems can achieve the same level of control as conventional systems that use a mouse cursor as the pointing device. Lutteroth et al. designed another novel dwell-based web browser called Actigaze [10]. Hyperlinks close to the current gaze position were labeled with colors, and could be selected by the user dwelling on a confirm button with the corresponding color at the side of the screen.

Other user interfaces monitor the users’ natural eye movements subtly in the background to infer their intent, and then offer appropriate assistance. There are known as attentive user interfaces [2]. Cig and Sezgin used natural eye movements to improve real-time recognition of manipulation commands in pen-based device using a hidden Markov model (HMM) [11]. Dong et al. developed a hybrid gaze/electroencephalography (EEG) interface which suppressed the selection of unlikely commands depending on the user’s intent estimated from their natural gaze trajectories [12, 13]. In the field of human-robot interaction, Das et al. designed a robot which was able to determine the suitable time to interact with a person by estimating the person’s visual focus of attention based on the cues including gaze patterns [14]. Huang and Mutlu studied eye gaze in a scenario where a robot delivered an item to a user who ordered the item from a menu using speech. The robot anticipated the user’s intent from his/her gaze patterns before the speech command using a support vector machine (SVM), and then planned its motion in advance or even performed anticipatory actions. This anticipatory robot system increased task completion speed [15]. Gaze trajectories have also been used to improve web browsing performance. Rozado et al. increased the web navigation speed by prefetching the websites corresponding to hyperlinks upon which the users had fixated.

This work was supported in part by the Hong Kong Research Grants Council General Research Fund grant #618713.

The authors are with the Department of Electronic and Computer Engineering (ZKC, BES) and the Division of Biomedical Engineering (BES) at the Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong. (e-mail: zhchenbc@ust.hk, eebert@ust.hk).
for more than 300ms [16], Lee et al. reduced the initial delay in web video access by prefetching the video based on the click possibility calculated from the cursor-gaze movement relationship [17].

The choice of this dwell time is a tradeoff between speed and the rate of unintentional selection: the shorter the dwell time, the faster objects can be selected, but the higher the rate of unintentional selection. Most systems allowed users to adjust this dwell time to suit their personal habits. However, past work using dwell time, including all those described above, has considered only the use of a fixed dwell time applied uniformly across all objects of potential interest.

In this paper, we propose a method by which the dwell time can be adjusted dynamically and non-uniformly for different objects depending upon estimates of the probabilities that the user wishes to select them, and apply this method to improve web browsing speed in a gaze-based web browser whose user interface is shown in Fig. 1. Similar to Windows Control and GazeTheWeb, our system utilized a two-step process control policy for hyperlink selection. If users were interested in following a link, they first gazed at the “Select” button on the right hand side of the screen, and then selected the hyperlink they wished to follow by gazing at it. However, the dwell times assigned to different hyperlinks varied.

This paper makes three main contributions. First, we develop a new probabilistic model of eye gaze behavior during web surfing. This model enables us to estimate the probability that the user wishes to select each hyperlink on the web page based on his/her natural gaze behavior. Second, we propose a variable dwell time assignment policy, which sets the dwell time associated with each hyperlink depending upon its probability of being the target. Links that are more likely to be of interest have shorter dwell times to increase response speed, whereas less likely links have longer dwell times to avoid unintentional selection. Third, we provide the specific guidance in choosing the parameters of the dwell time selection policy so as to achieve the best tradeoff between selection speed and false selection. Our experimental results demonstrate that this method can achieve better tradeoff between speed and accuracy, and that there was a significant improvement compared to the ones using uniform dwell time. Because the basic paradigm for selection is unchanged, our variable dwell time algorithm does not increase cognitive load on the users.

II. GAZE-BASED WEB BROWSER INTERFACE

The graphical user interface of the gaze-based web browser used in this study is presented in Fig. 1. It is written in JavaScript as a Chrome Extension. It can communicate with MATLAB via a Chrome App using TCP/IP protocol. When a new page is loaded or changes happen on current web page, the user interface captures the locations and sizes of the bounding box of all hyperlinks on current page and sends them to MATLAB, which implements the probabilistic model described in next section.

This user interface provides totally four command buttons, “Back”, “Select”, “Cancel” and “Forward”, placed in the task bar fixed on the right hand side of the web browser. Commands are activated if the user maintains his/her gaze on the corresponding button for a fixed dwell time of 400ms. Once activated, the button color turns red. The “Back” (“Forward”) button is used to go backward (forward) in the browsing history, as with normal web browsers. The “Select” button is used to start the selection phase, during which a hyperlink can be selected by maintaining gaze on the desired hyperlink. The “Cancel” button is used to cancel the most recent command or hyperlink selection. Users are not given feedback about the system’s estimate of their gaze location, as this may be distracting [5].

During the selection phase, a hyperlink \( m \) is selected if \( N_m \) of the most recent 1.5\( N_m \) gaze points are assigned to that hyperlink. We define \( T_m = N_m T_s \) to be the dwell time of hyperlink \( m \), where \( T_s = 16.67 \text{ms} \) is the sampling period of the gaze tracker. The larger window over which gaze points can be accumulated is commonly used to make gaze-based systems more tolerant of noise and jitter during fixation. Each gaze point, \( G = (x_g, y_g) \), is assigned to the closest hyperlink \( L \) according to the distance measure

\[
D(G,m) = \min_{(x,y) \in B_m} \left\{ \max \left| x-x \right|, \left| y-y \right| \right\} 
\]

where \( B_m \) is the bounding-box of hyperlink \( m \), as long as this distance is smaller than or equal to 40 pixels. If the distance is larger than 40 pixels, then \( G \) is not assigned to any hyperlink.

Once a hyperlink is selected, it changes its text color to red. This visual feedback allows the users to realize the moment when a hyperlink is selected and to realize which hyperlink is selected.

III. VARIABLE DWELL TIME ASSIGNMENT ALGORITHM

This section describes our probabilistic method to assign different dwell times to different hyperlinks. Our algorithm is based upon the hypothesis that the gaze trajectory of the user before activating the “Select” button can be used to infer the information about the user’s intended selection. Intuitively,
more recently viewed hyperlinks are more likely to be the object of interest. Thus, our variable dwell time assignment algorithm first uses a probabilistic model of the gaze trajectory to infer the probability that each hyperlink is the intended target. It then assigns shorter dwell times to more likely hyperlinks.

A. Target Inference based on Gaze Trajectories

We assign probabilities to different hyperlinks using a two-stage algorithm. The first stage converts the raw eye gaze trajectory into a sequence of fixations separated by saccades, known as the scanpath [18]. The second stage estimates the probability that each hyperlink is the intended target based on the scanpath.

1) The First Stage Model

We model reading behavior as a sequence of fixation-saccade cycles. Accurate identification of fixations is important because the visual information is obtained during fixations [19]. The location and duration of fixations usually indicate the points of interest of the person. Thus, we propose an algorithm which automatically identifies fixations from sampled gaze data from eye tracker.

In contrast to algorithms which require choosing threshold parameters manually, e.g., velocity-threshold identification (I-VT) and dispersion-threshold identification (I-DT) [20], we use here a data driven algorithm. We extend the hidden Markov model identification (I-HMM) [20], which utilized a first-order autoregressive first-order HMM to label each gaze point as fixation or saccade, to a second-order autoregressive second-order hidden Markov model. The I-HMM was vulnerable to noise because it only considered first-order dependency. Introducing a higher order dependency makes our model more robust to variation and noise, and enables us to address the problem of single outlier during fixations [21]. We expect that our model can achieve better gaze behavior modelling and thus improve target link inference. We denote our model I-HMM2.

Our I-HMM2 labels each gaze point as a fixation, saccade or outlier. The evolution of joint probability of sequence of hidden states \(l_1; l_2; \ldots; l_T\) and sequence of observations (gaze points, \(g_1, g_2, \ldots; g_T\)) is given by:

\[
p(l_1; \ldots; l_T, g_1; g_2; \ldots; g_T) = p(l_1, g_1) \prod_{t=2}^{T} p(l_t | l_{t-1}, g_{t-1}) \prod_{i=2}^{T} p(g_i | l_i, \ldots, l_1, g_{i-1}, \ldots, g_2)
\]

where \(l_t \in \{\text{fixation (f), saccade (s), outlier (o)}\}\) is the hidden label of \(g_t\). Transitions between state \(o\) and state \(s\) are not allowed, as we assume that the outliers only occur during fixations. Self-transition of state \(o\) is not allowed either for the same reason.

The emission distributions, \(p(g_t | l_1; l_2; \ldots; l_{t-1}; g_{t-1})\), are assumed to be two-dimensional Gaussian distributions, i.e.,

\[g_t \sim N(\mu_g (l_{t-2} \ldots l_1; g_{t-1}), \Sigma_g (l_{t-2} \ldots l_1; l_t))\]

where \(\mu_g\) and \(\Sigma_g\) are given by:

\[
\mu_g = \begin{cases} 
\frac{1}{2} (g_{t-2} + g_{t-1}) & l_{t-2} \neq o, l_{t-1} = f \\
g_{t-2} & l_{t-2} = f, l_{t-1} = o \\
g_{t-1} & l_{t-2} = o, l_{t-1} = f \text{ or } l_{t-2} \neq o, l_{t-1} = s
\end{cases}
\]

\[
\Sigma_g = \begin{cases} 
1.5 \Sigma_f & l_{t-2} \neq o, l_{t-1} = f, l_t = f \\
2 \Sigma_f & l_{t-2} = o, l_{t-1} = f, l_t = f \\
\Sigma_s & l_t = s \\
\Sigma_o & l_t = o
\end{cases}
\]

The initial probability, \(p(l_1|g_1; g_2)\), is assumed to be uniformly distributed over hidden state \(l_1\) and screen coordinate \((g_1; g_2)\). This setting has little impact on the performance of the model as it runs at 60Hz (the sampling rate of the gaze tracker) with long sequence of data.

Given a sequence of gaze points, \(g_1; g_2; \ldots; g_T\), we can use the Viterbi algorithm [22] to compute optimal sequence of labels. The parameters of this first stage model can be trained from unlabeled data by modifying the Baum-Welch algorithm [22]. Note that only the number of parameters of transition matrix increases (from six to 14) when extending a first-order model to a second-order model.

After the labels of every gaze data are defined, we then convert the raw eye gaze data into sequence of fixations. We first group the gaze data separated by two labeled saccades into a single fixation with an associated spatial coordinate on the screen \((x, y)\), which is obtained by averaging the locations only of gaze data labeled as fixations, and an associated duration \(d\), which is the time between the two saccade boundaries. Fixations with durations less than 100ms are discarded, since fixations are rarely shorter than 100ms [20]. The sequence of fixations is the scanpath.

2) The Second Stage Model

Intuitively, one might expect choice and gaze to be correlated. Indeed, when users were instructed to choose the more attractive between two faces, Shimojo et al. found that gaze was biased towards the face eventually chosen, especially in the seconds right before making decision [23].

In our case, we might expect the hyperlink that the user wants to select to be among the more recent fixations before the user activates the “Select” button. To evaluate the extent to which this intuition holds, we analyzed eye gaze data from Experiment I described below, where users freely browsed the web through the gaze browser. Fig. 2 shows that about 80% of the time the selected hyperlink was among the last three
The gaze behavior, $B_n$, assumes one of three integer values: 1 (reading away from the hyperlink), 2 (reading near the hyperlink), and 3 (reading away from the hyperlink). Both latent variables evolve according to independent Markov chains. The independence assumption enables the combined state transition matrix to be factorized, reducing the number of parameters [24].

Fig. 3 presents the FHMM model as a Bayesian network. The two latent variables are the intended target link for selection, $I_n$, and the gaze behavior, $B_n$. The intended target $I_n$ assumes integer values from 1 to $M$, the number of hyperlinks on the current page. The gaze behavior, $B_n$, assumes one of three integer values: 1 (reading away the hyperlink), 2 (reading near the hyperlink), and 3 (reading away from the hyperlink). Both latent variables evolve according to independent Markov chains which end at a terminal state (TS), corresponding to the activation of the “Select” button. The location and duration of the $t$th fixation in the sequence, $f_t$, depends on both $I_t$ and $B_t$. We consider both location and duration because longer fixations before the select action was chosen. The spatial location and the duration of the fixation are assumed to be conditionally independent. Thus, the emission distribution, $p(f_t | I_t, B_t)$, factorizes as

$$p(f_t = [x, y, d] | I_t, B_t) = p(x, y | I_t, B_t) \times p(d | B_t)$$

where $p(x, y | I_t = m, B_t = k)$ is given by a normal distribution with mean $\mu_m$ and covariance matrix $\Sigma_{mk}$ for $k = 1, 2$ and is uniformly distributed over the entire screen for $k = 3$. The mean is given by the center of the bounding box of hyperlink $m$. The covariance matrix is given by

$$\Sigma_{mk} = [\beta_{mk} W_m^2 + \sigma_{mk}^2, 0, 0, (\beta_{mk} H_m^2 + \sigma_{mk}^2)]$$

where $W_m$ and $H_m$ are the width and height of the hyperlink, respectively. The $\beta$ and $\sigma$ parameters are learned, and are constrained to be larger for $k = 2$ than $k = 1$. The distribution of the duration is assumed to be lognormal with mean and variance parameters that vary according to the behavior.

Moreover, the initial probability distribution is given by

$$p(z_1) = p(I_1 = m, B_1 = k) = \frac{\pi_k}{M}$$

Once the “Select” command is activated, we compute the probability that hyperlink $m$ is the intended target according to

$$p(I_t = m | z_{t-1} = \text{TS}, f_{t-1} = f_1, \ldots, f_{t-1}) = \sum_{n,k} p(I_t = m, B_t = k = 1, f_1, \ldots, f_{t-1}) p(z_{t-1} = \text{TS} | B_t = k)$$

The probability $p(I_t = m, B_t = k = 1, f_1, \ldots, f_{t-1})$ can be computed using the forward algorithm [22]. We have assumed that the transition to the terminal state depends only on the gaze behavior and not the target identity, i.e.,

$$p(z_{t+1} = \text{TS} | z_t) = p(z_{t+1} = \text{TS} | B_t)$$

To save time, we consider only at most the most recent $T=5$ fixations on the scanpath, based on a preliminary study (data not shown).

We trained the parameters of gaze behavior transition probabilities and the emission distributions in a semi-supervised manner. We used a slightly modified Baum-Welch algorithm, but assumed that the intended target $I_t$ is always equal to the known target for all $t$. These parameters were then fixed, and the optimal value of the $p_i$ parameter was
found by grid-search in a second pass over the training data while \( I_c \) is not constrained.

This FHMM model accounts for variability in the locations, sizes and number of hyperlinks on different web pages. It also allows for variability in the length of the scanpath, and accounts for variability in the gaze due to noise or jitter through the use of the Gaussian distribution.

### B. Variable Dwell Time Selection

We assign a different dwell time to each hyperlink using a dwell time selection policy, which typically decreases with the probability that the hyperlink is the intended target. In other words, the nominal dwell time of hyperlink \( m \) is given by

\[
T_m = h(p(I = m | z_{T+1} = TS, f_1 \cdots f_r))
\]

(13)

where \( h() \) is a non-increasing function. Due to sampling by the gaze tracker, the actual dwell time is obtained by quantizing nominal dwell time to an integer number of samples.

Here, we choose \( h() \) among a set of piecewise linear policies determined by four parameters \([T_{max}, T_{min}, T_{break}, p_{break}]\),

\[
h(p) = \begin{cases} T_{max} - (T_{max} - T_{break}) \frac{p}{p_{break}} & \text{if } 0 \leq p \leq p_{break} \\
T_{break} - (T_{break} - T_{min}) \frac{p - p_{break}}{1 - p_{break}} & \text{if } p_{break} < p \leq 1 \end{cases}
\]

(14)

This function drops linearly from \( T_{max} \) at \( p = 0 \) to \( T_{break} \) at \( p = p_{break} \) and then to \( T_{min} \) at \( p = 1 \). To ensure the function is non-increasing, we constrain \( 0 < T_{min} \leq T_{break} \leq T_{max} \).

Fig. 4 shows two example policies.

### IV. EXPERIMENTAL PROCEDURES

#### A. Setup

During experiments, subjects were seated about 65 cm in front of a 19 inch monitor with 1280 \( \times \) 1024 pixel resolution. A chin-rest was used to keep the positions of their heads stable. Eye gaze data were recorded at 60 Hz by a Tobii X60 eye tracker mounted under the monitor. Before each experiment, the eye tracker was calibrated using the standard nine-point calibration, followed by a validation session to check whether the eye tracker can record the gaze points accurately enough to distinguish two adjacent rows on the web pages. If not, calibration was repeated. Calibration was also repeated during breaks if necessary.

Before each experiment, subjects were given a practice session to familiarize themselves with the gaze-based browser.

#### B. Participants

In total, 25 subjects (16 males and 9 females) participated in this study. All had normal or corrected-to-normal vision. None had prior experience with gaze-based web browsing. We conducted three experiments, but as shown in Table I, not all subjects participated in all three experiments. The experimental procedures involving human subjects described in this paper were approved by Committee on Research Practices at the Hong Kong University of Science and Technology. All subjects provided written consent.

#### C. Experiment I

In this experiment, subjects were free to browse the web using the gaze browser. The dwell times of hyperlinks were all set to 500 ms. Fourteen subjects (nine males and five females) participated in this experiment. Each subject performed one or two sessions depending on their self-reported fatigue level, with a break between each session.

Each session of the experiment starting by presenting the user with a webpage of his/her choice, and then allowing him/her to choose and follow hyperlinks freely. The session ended after the user had followed five to six hyperlinks.

If a link was incorrectly selected by the system, users were instructed to either cancel the selection and try again or to report the mis-selection and the intended target link to the experimenter. This instruction allowed us to obtain the intended target links of the users.

Each session was divided into trials, where each trial corresponds to the gaze trajectory from the time a web page is first presented until the first attempt at selecting a hyperlink is made, irrespective of whether that attempt resulted in a correct or an incorrect selection by the system. On average, one trial lasted for about 30 s containing 78 fixations.

In total, we collected gaze data from 86 valid trials. Five trials were considered invalid because of significant eye tracking inaccuracy.

#### D. Experiment II

While free viewing is a more natural task, its drawback is that it results in relatively few selections per unit time, as
subjects spend most of their time reading the text of the webpage. To accelerate data collection and provide for better experimental control, subjects in Experiments II were asked to perform a search-and-select task. Twelve subjects (nine males and three females) participated.

In each trial, the subjects were asked to find and select a particular hyperlink. The text of the desired hyperlink was presented to the user at the beginning of each trial, and also presented at the top of the screen during the trial in case the subjects forgot. A trial ended when a hyperlink was selected, regardless of the correctness. A trial was considered valid if the selection was made within 45s from the start of the trial and if a hyperlink was selected within 2s after “Select” button was activated.

Each subject participated in four sessions, each containing 10 trials. Subjects were given a break between the second and third session. The sessions differed according to the dwell time selection policy used and the set of web pages/desired hyperlinks presented to the users.

We collected data for four different dwell time selection policies: three where the dwell times for all hyperlinks were the same (100ms, 300ms and 500ms) and one where the dwell time varied according to our proposed algorithm with parameters [600ms, 100ms, 100ms, 1], i.e., the nominal dwell time decreased linearly from 600ms to 100ms. The actual dwell time was quantized to a multiple of 100ms, i.e., \( N_d = \lceil \frac{T_m}{60}\rceil \), resulting a range from 500ms to 100ms. All subjects used each policy once, but in random order. Subjects were told that the policies would vary between sessions, but were not told which policy they were using in each session.

We pre-defined four sets of web pages/desired hyperlinks to balance the difficulty among different sessions and among different subjects for better experimental control. Desired hyperlinks were chosen from more cluttered locations on the webpages. Each session used a different set, but the order in which the sets were presented was randomized independently of the policy order. The parameters of the gaze model were trained on all data from Experiment I.

After each session, we asked the subjects to evaluate accuracy/response speed of the different policies using a 5-point Likert-like scale, where 1 means very inaccurate/slow and 5 means very accurate/fast. We also asked subjects to rank the policies based on their overall preference. The subjects could revise their scores/ranking at the end of experiment to ensure consistency.

In total, we collected data from 477 valid trials. One trial using the 100ms dwell time policy and two trials using the 300ms dwell time policy were discarded as invalid.

E. Experiment III

Experiment III was similar to Experiment II, except that we collected data from six different dwell time policies over six sessions:

- Uniform: 500ms
- Variable: [500ms, 16.67ms, 50ms, 0.005]
- Variable: [500ms, 16.67ms, 50ms, 0.3]
- Variable: [500ms, 16.67ms, 16.67ms, 1]
- Variable: [500ms, 100ms, 100ms, 1]
- Variable: [500ms, 316.67ms, 316.67ms, 1]

The rationale behind these choices is described in more detail below. The parameters of the gaze model were also trained on the data from Experiment I.

Ten subjects (seven males and three females) participated in this experiment. We collected data from 598 valid trials. One trial under the [500ms, 16.67ms, 50ms, 0.3] policy and one trial under the [500ms, 16.67ms, 16.67ms, 1] policy were invalid. Subjects were not asked to evaluate the different policies.

V. EXPERIMENTAL RESULTS

A. Target Inference from Gaze Trajectories

We used the data from Experiment I to train and test the eye gaze model. All testing results were computed using fourteen-fold leave-one-subject-out cross-validation. Model parameters reported here were obtained by averaging the model parameters across the fourteen folds.

The average parameter values for the second state model after training are illustrated in Fig. 5 and given explicitly in Table II and Table III. Consistent with the finding that long fixations are usually related to user’s interest [25], we find that the mean of the distribution of fixation duration when the user...
The performance of \( B = 1 \) to be larger than the mean of other distributions (Fig. 5 (b)).

We evaluated the model by testing its accuracy in inferring the hyperlink the user wished to select, which we assumed to be the link \( m \) with the highest \( p(I = m | z_{t-1} = \text{TS}, f_1 \cdots f_T) \). The classification accuracy obtained from fourteen-fold cross-validation was 65.12%. Web pages in the testing set contained 53.30 hyperlinks on average, implying chance accuracy of 1.88%. This classification accuracy is also higher than that obtained by the simple heuristic of choosing the last hyperlink fixated upon by the user (56%, Fig. 2).

For a more nuanced characterization of the model, Fig. 6 shows histograms of the probabilities \( p(I = m | z_{t-1} = \text{TS}, f_1 \cdots f_T) \) assigned to the true target link, the highest probability among the non-target links, and the difference between them. About 50% of the probabilities of the true target link were greater than 0.8, whereas over 50% of the probabilities of the most likely non-target link were smaller than 0.2. Although the model assigned a probability less than 0.2 to the true target about 30% of the time, it assigned a probability greater than 0.8 to the most likely non-target link about 10% of the time. This suggests that when the true target link is assigned a low probability, most of the other hyperlinks also have low probabilities, i.e., the model is unsure. This is consistent with the histogram of the differences: about 35% of the differences were greater than 0.8, whereas only about 10% of them were lower than -0.8.

Moreover, to compare our proposed I-HMM2 with I-HMM, we tested on the data from all three experiments. To test on the data from Experimental I, we retrained and retested the whole gaze model but using I-HMM as the first stage model in the same way we trained and tested the gaze model with I-HMM2 described above, i.e., by cross-validation. To test on the data from Experimental II and III, we trained both models on the data only from Experiment I. The average target inference accuracy of the simple heuristic of choosing the last hyperlink fixated upon, the gaze model with I-HMM and the gaze model with I-HMM2 are reported in Table IV. The results show that the gaze models with I-HMM2 achieve about 4.5% improvement compared to that with I-HMM. This improvement resulted from replacing I-HMM with I-HMM2 implies better modeling of gaze by I-HMM2. Moreover, it is worth noting that both gaze models achieve higher average inference accuracy than the simple heuristic for all three experiments (on average, about 6.3% gain for the gaze model with I-HMM and about 10.8% gain for the gaze model with I-HMM2).

### Table II
Average Transition Matrix for Gaze Behavior of the Second Stage Model

| \( B_1 \) | \( B_2 \) | \( B_3 \) | TS |
|---|---|---|---|
| 1 | 0.57 | 0.00 | 0.08 | 0.35 |
| 2 | 0.34 | 0.55 | 0.03 | 0.08 |
| 3 | 0.05 | 0.16 | 0.59 | 0.20 |

### Table III
Average Trained Parameters of the Emission Distributions of the Second Stage Model

| \( \beta_{11} \) | \( \beta_{12} \) | \( \beta_{13} \) | \( \mu_{I} \) | \( \mu_{II} \) | \( \mu_{III} \) | \( \sigma_{I} \) | \( \sigma_{II} \) | \( \sigma_{III} \) |
|---|---|---|---|---|---|---|---|---|
| 0.17 | 0.48 | 0.00 | 2.90 | 2.64 | 2.54 | 0.08 | 0.66 |
| \( \sigma_{I} \) (px) \( \sigma_{II} \) (px) \( \sigma_{III} \) (px) | \( \sigma_{I} \) | \( \sigma_{II} \) | \( \sigma_{III} \) | \( \sigma_{I} \) |
| 39.38 | 14.07 | 126.43 | 38.41 | 0.62 | 0.45 | 0.47 | 0.26 |

* These values were smaller than 5e-3.

### Table IV
Average Target Inference Accuracy of Three Experiments by Three Methods

| | Heuristic | Gaze model with I-HMM | Gaze model with I-HMM2 |
|---|---|---|---|
| Experiment I | 55.8% | 60.5% | 65.1% |
| Experiment II | 66.9% | 74.2% | 77.4% |
| Experiment III | 68.9% | 75.9% | 81.6% |

B. Variable Dwell Time Policies: Simulation

Using the data collected in Experiments I, II and III when users were using the uniform 500ms dwell time policy, we simulated the performance of different dwell time policies. We did this by using the gaze trajectory before the user activated the “Select” button to estimate the probability that each link was the target link using the probabilistic gaze model, assigning different dwell times to different hyperlinks based on these probabilities, and then predicting which hyperlink would be selected based on the gaze trajectory after the user activated the “Select” button.

This enabled us to predict the performance of a wide range of different dwell time selection policies without extensive experimentation. We were hopeful that these simulations would give a good indication of the performance of these policies during actual use. Since visual feedback was only provided at the moment either a command button or a hyperlink was selected, we assumed that the actual dwell time selection policy being used would not significantly influence the gaze behavior if the maximum dwell time \( T_{\text{max}} \) was smaller than the 500ms used in the experimental data. The next section compares our simulated results with our experimental results.

We used grid search to perform exhaustive simulation over the entire range of non-decreasing dwell time selection policies such that \( 0 < T_{\text{min}} \leq T_{\text{max}} \leq 500 \text{ms} \). In particular, we varied \( T_{\text{min}} \) from 16.66ms to 500ms, \( T_{\text{max}} \) from 16.66ms to \( T_{\text{max}} \), \( T_{\text{break}} \) from \( T_{\text{min}} \) to \( T_{\text{max}} \), and \( p_{\text{break}} \) from 0 to 1. The grid step size was 16.67ms (1 sample) in time and 0.1 in probability. Note that this search space includes the set of all uniform dwell time policies, which are obtained by setting \( T_{\text{min}} = T_{\text{max}} \).

For each policy, we evaluated the average error rate and average response time. The error rate is defined as the percentage of incorrect hyperlink selections. The response time is defined as the duration between the moment when the user moves his/her eyes away from the “Select” button after activating it and the moment when a hyperlink is selected.

Fig. 7 presents the simulation results and a comparison with experimental results, which will be discussed in more detail in next section. For the simulation results, the performance of each policy is shown as a cyan marker. The commonly used uniform dwell time policy is shown as a red curve. The curve
does not apply.
In Experiment II, subjects were asked to rate each policy according to the accuracy and response speed and rank the policies according to overall preference. Scores of accuracy and response speed given by each individual subject were mean-centered by subtracting the mean scores of the subject. Fig. 9 shows the mean-centered scores of accuracy and response speed. The directions of the axes were reversed so that the plot is comparable to Fig. 7, i.e. lower accuracy/higher error rate in the upper part of the graph and slower speed/longer response time on the right hand side. The average scores of the 100ms policy, 300ms policy, 500ms policy and variable dwell time policy, in terms of (score of speed, score of accuracy) are (1.10, -0.88), (-0.23, -0.21), (-1.06, 0.46) and (0.19, 0.63), respectively. Similar to the objective results, the variable dwell time policy lies to the lower left of the line connecting the uniform dwell time policies, indicating that subjects felt that it achieved a better tradeoff between speed and accuracy. Note that the order of scores of accuracy is consistent to the order of objective response time, indicating the differences of speed were perceived by the subjects.

Statistical tests indicated that the observed differences were significant. A repeated measures ANOVA on the response speed scores indicated a statistically significant effect of dwell time policies ($F(3, 33) = 27.91, p < .0001$). Post-hoc tests indicated significant differences in response speed between the variable dwell time policy and the 100ms ($p < .01$) and 500ms ($p < .01$) policies, but not between the variable dwell time policy and the 300ms policy ($p = .42$). Similarly, a repeated measures ANOVA on the accuracy scores indicated a statistically significant effect of dwell time policies ($F(3, 33) = 12.76, p < .0001$). Post-hoc test indicated significant differences between the variable dwell time policy and the 100ms ($p < .001$) and 300ms ($p < .05$) policies, but not between the variable dwell time policy and the 500ms policy ($p = .95$).

Fig. 10 presents the ranking results of different dwell time policies. Consistent with previous results, subjects tended to prefer the variable dwell time policy. The mean and standard deviations of the 100ms, 300ms, 500ms and variable policy were $3.33 \pm 0.98$, $2.58 \pm 1.08$, $2.58 \pm 1.00$ and $1.50 \pm 0.67$, respectively. Under two-tailed Wilcoxon signed-rank tests, the ranking of variable dwell time policy was statistically significantly different from the ranking of the three uniform dwell time policies (100ms, $Z = 2.87, p < .01$; 300ms, $Z = 2.17, p = .030$; 500ms, $Z = 2.18, p = .029$).

### VI. DISCUSSION AND CONCLUSION

Past work in gaze-based selection has typically used a single dwell time applied uniformly across all objects to avoid the “Midas touch” problem. The choice of this dwell time is a tradeoff between response speed and error rate. To achieve a better tradeoff, we propose here a method for varying the dwell time between different objects based on the probability that the user wishes to select the object (user intent), which can be inferred from prior gaze behavior.

In order to infer user intent, we proposed a two-stage probabilistic model for natural gaze trajectories and trained it on data collected from users as they surfed the web using a gaze-based web browser using a uniform dwell time for selection of 500ms.

In the first stage, the model segmented the past gaze trajectory into a sequence of fixations. In the second stage, the model used this sequence to infer the probability of each hyperlink on the page being the one the user would like to select. When trained on natural gaze behavior, the model is consistent with experimental findings about gaze and choice [23] [25]. This model was subject-independent. Our experiments with leave-one-subject-out cross validation show that the model can identify the user’s target hyperlink with accuracy much better than chance and better than a simple heuristic, which says that the user probably wants to select the last item s/he looked at.

---

**Table V**

| Error Rate | Response Time (ms) | Error Rate | Response Time (ms) |
|------------|--------------------|------------|--------------------|
| 100ms uniform policy | 32.77% ± 8.47% | 168.77 ± 5.74 | 32.03% ± 5.24 | 178.27 ± 8.66 |
| 300ms uniform policy | 21.19% ± 7.40% | 508.47 ± 33.59 | 16.67% ± 4.18 | 473.31 ± 15.75 |
| [600ms, 100ms, 100ms, 1] | 10.00% ± 5.39% | 372.08 ± 25.72 | 12.42% ± 3.70 | 388.29 ± 29.82 |
| 500ms uniform policy | 13.33% ± 6.11% | 730.83 ± 38.35 | 11.11% ± 3.53 | 733.55 ± 22.78 |

* Each entry is presented as Mean ± 95% CI (confidence interval)

**Table VI**

| Error Rate | Response Time (ms) | Error Rate | Response Time (ms) |
|------------|--------------------|------------|--------------------|
| 500ms, 16.67ms, 500ms, 0.0005 | 30.00% ± 9.03% | 72.33 ± 10.68 | 28.43% ± 5.06 | 90.34 ± 13.71 |
| 500ms, 16.67ms, 500ms, 0.3 | 23.23% ± 8.36% | 109.26 ± 21.99 | 18.63% ± 4.37 | 147.93 ± 20.89 |
| 500ms, 16.67ms, 16.67ms, 1 | 16.16% ± 7.29% | 198.82 ± 38.35 | 13.40% ± 3.82 | 274.24 ± 27.94 |
| 500ms, 100ms, 100ms, 1 | 11.00% ± 6.16% | 304.33 ± 34.34 | 13.07% ± 3.78 | 385.51 ± 27.63 |
| 500ms, 316.7ms, 316.7ms, 1 | 7.00% ± 5.03% | 524.00 ± 40.63 | 9.48% ± 3.29 | 594.06 ± 24.31 |
| 500ms uniform policy | 13.00% ± 6.62% | 728.83 ± 38.69 | 11.11% ± 3.53 | 733.55 ± 22.78 |

* Each entry is presented as Mean ± 95% CI (confidence interval)
Based on the inferred probabilities about user intent, we proposed an algorithm to assign variable dwell times to different links, with more likely links having shorter dwell times. In particular, we considered a four-parameter class of policies where the dwell time of a link decreased in a piecewise linear manner with the inferred probability that the link is the user’s desired selection. We performed extensive simulations to identify the policy parameters that resulted in the best tradeoff between accuracy and response speed, and identified two single parameter policies which could achieve the best tradeoff.

Our experimental results evaluating these policies both objectively (Fig. 7, Table V and Table VI) and subjectively by users (Fig. 9 and Fig. 10), demonstrate that our proposed method can achieve a better tradeoff between selection time and accuracy than uniform dwell times. Note that because the basic paradigm for selection is unchanged, our proposed method does not increase cognitive load on the users.

Moving forward, we plan to extend this method by taking more information into account in the intent inference model, e.g., the information of the users’ previous command [26] and information about the current task. Because we have formulated our model probabilistically, we anticipate that with the proper formulation, we will be able to incorporate this information in a mathematically well founded and rigorous manner.

In closing, we would like to note that the basic framework we have used here, i.e. two-step dwell-based selection is commonly used in current state-of-the-art systems, like Windows Control and GazeTheWeb. Thus, our algorithm is totally compatible with these kind of systems, and could be adopted with minimal change to existing interfaces and with little to no retraining of users. The technique is also complementary to other approaches for improving gaze-based interfaces, such as fisheye lenses [27], automatic magnification [8] [9], and real-time visual feedback [28]. Thus, we expect that the combination of our algorithm and these techniques can yield more adaptive and stable eye-based systems with better performance.

REFERENCES

[1] A. Poole and L. J. Ball, "Eye tracking in human-computer interaction and usability research: current status and future prospects," Encyclopedia of Human Computer Interaction, vol. 1, pp. 211-219, 2006.
[2] P. Majaranta and A. Bulling, "Eye tracking and eye-based human–computer interaction," in Advances in Physiological Computing, ed: Springer, 2014, pp. 39-65.
[3] T. O. Zander, M. Gaertner, C. Kothe, and R. Vilímek, "Combining eye gaze input with a brain–computer interface for touchless human–computer interaction," International Journal of Human–Computer Interaction, vol. 27, pp. 38-51, 2010.
[4] P. Majaranta, U.-K. Ahola, and O. Špakov, "Fast gaze typing with an adjustable dwell time," in Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, 2009, pp. 357-360.
[5] R. J. Jacob, "Eye tracking in advanced interface design," Virtual Environments and Advanced Interface Design, pp. 258-288, 1995.
[6] K.-J. Rääihä and S. Ovaska, "An exploratory study of eye typing fundamentals: dwell time, text entry rate, errors, and workload," in Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, 2012, pp. 3001-3010.
[7] A. Murata, "Eye – gaze input versus mouse: Cursor control as a function of age," International Journal of Human – Computer Interaction, vol. 21, pp. 1-14, 2006.
[8] Windows Control-gaze enabled computer access. Available: http://www.tobiiidynavox.com/windows-control/
[9] R. Menges, C. Kumar, K. Sengupta, and S. Staub, "eyeGUI: A Novel Framework for Eye-Controlled User Interfaces," in Proceedings of the 9th Nordic Conference on Human-Computer Interaction, 2016, p. 121.
[10] C. Lutteroth, M. Penkar, and G. Weber, "Gaze vs. mouse: A fast and accurate gaze-only click alternative," in Proceedings of the 28th Annual ACM Symposium on User Interface Software and Technology, 2015, pp. 385-394.
[11] Ç. Çığ and T. M. Sezgin, "Gaze-based real-time activity recognition for proactive interfaces," in Proceedings of the 23rd Signal Processing and Communications Applications Conference, 2015, pp. 694-697.
[12] X. Dong, H. Wang, Z. Chen, and B. E. Shi, "Hybrid Brain Computer Interface via Bayesian integration of EEG and eye gaze," in Proceedings of the 7th International IEEE/EMBS Conference on Neural Engineering, 2015, pp. 150-153.
H. Wang, X. Dong, Z. Chen, and B. E. Shi, "Hybrid gaze/EEG brain computer interface for robot arm control on a pick and place task," in Proceedings of the 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 2015, pp. 1476-1479.

D. Das, M. G. Rashed, Y. Kobayashi, and Y. Kuno, "Supporting Human-Robot Interaction Based on the Level of Visual Focus of Attention," IEEE Transactions on Human-Machine Systems, vol. 45, pp. 664-675, 2015.

C.-M. Huang and B. Mutlu, "Anticipatory robot control for efficient human-robot collaboration," in Proceedings of the 11th ACM/IEEE International Conference on Human-Robot Interaction, 2016, pp. 83-90.

D. Rozado, A. El Shoghri, and R. Jurdak, "Gaze dependant prefetching of web content to increase speed and comfort of web browsing," International Journal of Human-Computer Studies, vol. 78, pp. 31-42, 2015.

S. Lee, J. Yoo, and G. Han, "Gaze-assisted user intention prediction for initial delay reduction in web video access," Sensors, vol. 15, pp. 14679-14700, 2015.

D. Noton and L. Stark, "Scanpaths in saccadic eye movements while viewing and recognizing patterns," Vision Research, vol. 11, pp. 929-938, 1971.

K. Rayner, "Eye movements in reading and information processing: 20 years of research," Psychological Bulletin, vol. 124, p. 372, 1998.

D. D. Salvucci and J. H. Goldberg, "Identifying fixations and saccades in eye-tracking protocols," in Proceedings of the Symposium on Eye Tracking Research and Applications, 2000, pp. 71-78.

M. Kumar, "Gaze-enhanced user interface design," Ph.D. dissertation, Department of Computer Science, Stanford University, Stanford, CA, 2007.

L. R. Rabiner, "A tutorial on hidden Markov models and selected applications in speech recognition," Proceedings of the IEEE, vol. 77, pp. 257-286, 1989.

S. Shimojo, C. Simion, E. Shimojo, and C. Scheier, "Gaze bias both reflects and influences preference," Nature Neuroscience, vol. 6, pp. 1317-1322, 2003.

Z. Ghahramani and M. I. Jordan, "Factorial hidden Markov models," Machine Learning, vol. 29, pp. 245-273, 1997.

J. L. Orquin and S. M. Loose, "Attention and choice: A review on eye movements in decision making," Acta Psychologica, vol. 144, pp. 190-206, 2013.

D. D. Salvucci and J. R. Anderson, "Intelligent gaze-added interfaces," in Proceedings of the SIGCHI conference on Human Factors in Computing Systems, 2000, pp. 273-280.

M. Ashmore, A. T. Duchowski, and G. Shoemaker, "Efficient eye pointing with a fisheye lens," in Proceedings of Graphics Interface, 2005, pp. 203-210.

C. Kumar, R. Menges, and S. Staab, "Eye-Controlled Interfaces for Multimedia Interaction," IEEE MultiMedia, vol. 23, pp. 6-13, 2016.