IEICE TRANS. INF. & SYST., VOL.E93–D, NO.6 JUNE 2010

PAPER Special Section on Info-Plosion

Estimates of User Interest Using Timing Structures between Proactive Content-Display Updates and Eye Movements

Takatsugu HIRAYAMA[†a)], ***Member***, **Jean-Baptiste DODANE**[†], ***Nonmember***, **Hiroaki KAWASHIMA**[†], ***Member***, ***and*** **Takashi MATSUYAMA**[†], ***Fellow***

SUMMARY People are being inundated under enormous volumes of information and they often dither about making the right choices from these. Interactive user support by information service system such as concierge services will effectively assist such people. However, human-machine interaction still lacks naturalness and thoughtfulness despite the widespread utilization of intelligent systems. The system needs to estimate user's interest to improve the interaction and support the choices. We propose a novel approach to estimating the interest, which is based on the relationship between the dynamics of user's eye movements, i.e., the endogenous control mode of saccades, and machine's proactive presentations of visual contents. Under a specially-designed presentation phase to make the user express the endogenous saccades, we analyzed the timing structures between the saccades and the presentation events. We defined *resistance* as a novel time-delay feature representing the duration a user's gaze remains fixed on the previously presented content regardless of the next event. In experimental results obtained from 10 subjects, we confirmed that *resistance* is a good indicator for estimating the interest of most subjects (75% success in 28 experiments on 7 subjects). This demonstrated a higher accuracy than conventional estimates of interest based on gaze duration or frequency.

key words: *proactive interaction, Mind Probing, interest estimation, gaze behavior, time-delay feature*

1. Introduction

1.1 Mind Probing

In the information society, people can access enormous sources of data easily and search for their favorite information. They need to have a clear-choice criterion for large amounts of information to hit their target. If this is not clear, they will feel unsure and uneasy, that is, they will feel very stressed. In this situation, interactive support by a well-informed agent such as the concierge of a luxury hotel will effectively alleviate their concerns. Our research aims at creating a concierge system that can estimate interest of user and casually provide sensible information to him/her. Based on interaction with this system, he/she can hunt for something on his/her mind and come across a target that unconsciously satisfies his/her intent.

Over the past few years, machines have spread their presence throughout numerous aspects of our everyday lives. We humans frequently deal with them like we do with partners. However, humans still behave with these machines as users who have to express clear requests and specific commands to obtain desired results. Humans still regard them as passive and reactive objects. Human-machine interaction is hence not as thoughtful as human communication despite its evolutions in affective computing and other related domains.

To make machines' behavior closer to that of humans, our research is inspired from basic scenes of interaction that occur in everyday life. To smooth interaction with a partner, we humans probe our partner's mental state. In other words, humans naturally adopt proactive behaviors, such as by bringing up general topics or establishing eye contact, while showing expectations of a reply. The proactive approaches make the partner reveal his/her internal (mental) state through vocal intonations, facial expressions, gazes, and so on. As a consequence, people are able to make the interaction evolve smoothly toward a state that suits both parties. This is all the more true in situations where the concierge tries to provide information or a service to a hesitant visitor. The concierge should probe the visitor's mind casually and understand his/her needs by analyzing reactions.

We believe that machines should acquire this behavior of (1) proactively approaching the user and (2) estimating his/her internal state based on reactions to it, without waiting for any commands from the user to move on toward more efficient and adaptive human-machine interaction. We call this concept *Mind Probing*. In this work, we focus on the relationship between the dynamics of information services and user reactions. Recently, digital signage has evolved. A great deal of advertising information can be dynamically provided to people. Advertisers will increase their demands for estimating user interest in dynamic-information services. We propose a method of dynamically presenting contents and novel features to estimate which content a user is interested in, based on *Mind Probing*. This work is a first step toward achieving concierge-like interactive systems that provide more relevant information according to what the user is interested in. Figure 1 shows our prototype system.

1.2 From Mental States to Eye Movements

Although some researchers have recently considered interest to be an emotion [1], it is basically a mental state that causes people to focus their attention on something. We

Manuscript received August 31, 2009.
Manuscript revised December 28, 2009.

[†]The authors are with the Graduate School of Informatics, Kyoto University, Kyoto-shi, 606–8501 Japan.
a) E-mail: hirayama@vision.kuee.kyoto-u.ac.jp

DOI: 10.1587/transinf.E93.D.1470

Fig. 1 Prototype information service system with a large display and three cameras (circled in white).

are attracted to its intimate connection with the attention. Degrees of attention are measured mostly by (1) the effort that the person provides to focusing on his/her goal and (2) the resistance that the person offers in opposing any influences [2]. We use these measurements of attention to evaluate the user interest.

How can we measure attention? We focus on overt attention that humans express with their eye movements. Based on the eye-mind hypothesis [3], eyes are often a window into the mind. Humans turn their gaze (central fovea of the retina) to an object of interest to obtain detailed information on it. The gazing action has fast jerky movements. These are called *saccades* and are closely related to attention [4]. The saccades are programmed under two different control modes: exogenous (bottom-up and stimulus-driven, depending on the salience of objects in the visual field) and endogenous (top-down and goal-directed, depending on the volition of the person). We have to specifically extract the endogenous saccades for analysis, which express the overt attention occurring in top-down processes [5], [6]. To achieve this, we design a scenario for proactively presenting contents to separately trigger the exogenous and the endogenous saccades.

In the situation we have assumed, the information service system displays various contents on a screen and estimates the user's interest through his/her eye movements. We need a precise and reliable "bridge" that reveals the connection between eye movements and contents. We consider that this bridge is created by the dynamics between them. The user will respond to proactive and dynamic presentations of contents. The response patterns of eye movements are sure to reflect his/her interest. We especially focus on the timing structures between presentation events and gaze switches, i.e., saccades. Furthermore, we consider that the dynamics of humans are most likely to reveal their true natures as these are often unconscious and are difficult to simulate.

According to the previous discussion on measuring attention, we compute two indicators for time delays. We call them *reaction* and *resistance*.

- *Reaction* represents the response time to switch a gaze to the next presentation.
- *Resistance* represents the duration of a gaze fixing on the previously presented content regardless of the next event that occurs in a different part of his/her visual field.

We make a hypothesis that the delays for the proactive presentations of contents relate to the interest and test the hypothesis through various experiments.

1.3 Related Researches

Many researchers have proposed ways of estimating user's mental states. Picard et al. analyzed passively sensed physiological behaviors to recognize user's emotions [7]. Kapoor et al. used multiple physiological modalities probabilistically combined to classify three mental states during a game task [8]. Qvarfordt et al. analyzed the duration of user's gaze fixing on an object and the frequency of his/her gaze entering and leaving the object in order to estimate degree of interest in the object [9]. However, these researchers' studies were based on passive sensing. Later, Onishi et al. analyzed the timing structure of proactive *face-turning* behavior in human-human communication [10]. It would be worth checking if the results were similar to those in human-machine communication.

As far as we know, no previous researchers have analyzed the dynamics of eye movements in response to machine proactive behaviors to estimate user interest. We have proposed *Mind Probing*, which combines interest, eye movements, and proactive behaviors. In our preliminary work, we divided user's cognitive states into two phases called "input" and "evaluate" using a scenario for proactively presenting contents to estimate the interest. "Input" is a state where the user reads all information of contents and "evaluate" is another state where the user compares some contents. It is highly possible that the gaze behaviors when the contents are being evaluated have a stronger relation with the interest than when information is being input into the brain. To induce the user to the "input" phase, the contents were exclusively presented in turns. On the other hand, the "evaluate" phase was triggered by simultaneously redisplaying all contents. We focused on the duration and the frequency of user gazes during the "evaluate" phase. The main weakness was that these two phases were not clearly separated. The user often reread the information to remember it during the "evaluate" phase. Such behaviors were not closely related to the interest.

2. Proactive Presentation and Hypothesis

2.1 Situation Description

The machine side has a system that provides information with the purpose of helping a user. It is a visual pool of new contents. The user makes a choice from them under his/her knowledge constrains. They face each other and interact at a

distance of approximately 1 m. The machine proactively approaches the user by presenting various contents on a large display.

The user interacts with the machine by only using his/her eyes, acquires information through his/her sense of vision, and communicates with the system by means of eye movements. The system, on the other hand, uses two different devices to interact with the user, the large display to send information and cameras to receive user signals (eye movements). In fact, the user is not asked to talk or point to an area. As shown in Fig. 1, the screen of the display is divided into several areas. In each area, content that consists of several pages of articles is presented. That is, the system simultaneously provides several contents to the user. The system estimates which content the user is interested in.

As a scenario after the interest estimation, we assume the following interaction, though we do not realize it in this work. The system provides information relating to the content of interest or provides contents similar to it. While providing such information, the system repeats *Mind Probing* to more deeply and accurately understand the user interest[†].

2.2 Scenario

We design the scenario for proactively presenting contents to extract the endogenous saccades. The system proactively presents various contents based on the scenario which consists of two phases to separate the two control modes of saccades.

- The **glimpse phase**: when we humans make a choice from various contents, most of us often glance over them first. In the first phase, the articles of new content are displayed at a very fast update rate on an area to permit the user to passively elicit the glancing behavior. After the presentation, other new contents are displayed in the same way, one area after another. The presentation method can elicit the exogenous saccades from the user. This phase will also make him/her aware about what kind of contents are likely to be displayed in which part of the visual field. The user builds the cognitive map, which is a mental representation of spatial locations, in his/her mind [11]. The presented information is stored in the cognitive map. Its purpose is to connect the environmental information to an image that has been built inside the mind to achieve the user's goal [12]. This is literally the "mind's eye". The quick update of the glimpse phase is set to prevent the user from reading and understanding all the information and to "tease" his/her interest.
- The **observation phase**: after browsing quickly, humans observe some contents carefully. In the second phase, the same contents are displayed once again, in the same area as in the glimpse phase, but for a longer period. That is, the same articles are redisplayed at a slow update rate. This time, the user can fulfill his/her interest by sufficiently reading some contents that just "teased" him/her before. As he/she had already seen them in the glimpse phase, we believe that the saccades will be mainly endogenous, by referring to the cognitive map. The important detail is that the system updates the contents in random order. The user cannot guess which content will be updated. However, the user can pay attention to areas of interest thanks to the cognitive map. This random presentation allows him/her also to compare partial contents when they are updated in parallel.

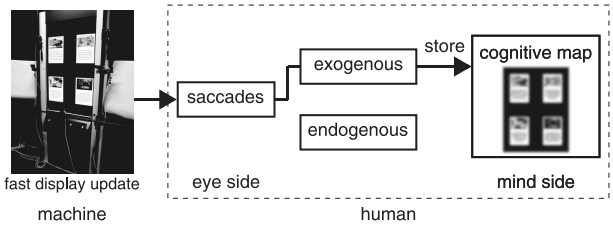

Fig. 2 Interaction scenario in glimpse phase.

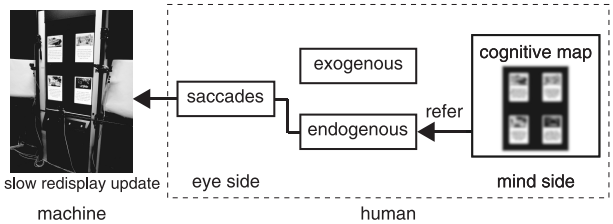

Fig. 3 Interaction scenario in observation phase.

Figures 2 and 3 have profiles of the scenario for proactively presenting contents.

2.3 Hypothesis: Dynamical Relation between Proactive Presentation Events and Eye Movements

We evaluate the user interest by calculating the delays between the presentations of contents and the gaze switches. In the framework of the proposed scenario, we make the following hypothesis:

- During the observation phase, we predict that the user will shift his/her gaze more quickly to the display update of interesting content, or he/she will continue to fix his/her gaze on the current content if it is more attractive than the next updated one. In other words, *reaction* will be shorter or *resistance* will be longer for the interesting content.

3. Apparatus and Software Architecture

3.1 System Overview

The proposed system remotely measures eye movements

[†]Indeed, we may need to combine the proposed method with a different type of proactive interaction such as using speech dialogue.

with no intrusive devices to achieve natural human-machine communication. It is composed of a large 50-inch display [†], three synchronized and calibrated cameras [††], and two backup lights, as shown in Fig. 1. The content articles are presented and controlled by a display software coded using WinAPI.

3.2 Gaze Estimation

The user gaze is estimated in four steps: face detection, estimation of face orientation, iris detection, and estimation of gaze direction.

First, the user's face is detected with the Intel OpenCV library, using Haar-like filters. Facial features (45 points) are extracted by using the Active Appearance Model (AAM) algorithm [13]. The AAM is a statistical subspace model of shape and appearance. The system has three AAMs for each user, i.e., an AAM for each user for each camera. Each AAM is trained using 15 images containing various rotations of the head in a preliminary experiment.

Then, a 3D face shape model is fitted onto the AAMs by the bundle adjustment [14]; in fact, the translation and rotation parameters are optimized using the steepest descent method. The 3D model consists of 45 feature points, eyeball centers, and iris radius, which are calibrated with the stereo cameras in the preliminary experiment. The 3D position of the user's face is estimated as a result of fitting. The irises are extracted by matching iris templates generated from the iris radius, and then their 3D positions are estimated from the eyeball centers and the iris radius.

After a straight line running through both the eyeball center and the iris center has been computed, its intersection with the display plane informs us about the gazing point. Finally, the estimation results from three cameras are integrated. The accuracy of gaze estimation is about 5 degrees (= about 10 cm on the screen).

4. Experimental Procedure

We conduct some experiments and estimate interest of several subjects by measuring the delays, the duration, and the frequency of their eye movements in response to the content presentation, and then evaluate the estimates by comparing the measurements with subjective results obtained from the subjects who are interviewed in a survey.

As recommended by Just and Carpenter [3], it is important to design a well-specified task in such experiments. This is because we want to observe and measure the endogenous saccades, which are goal driven. A goal therefore has to be defined for the subjects. In our experiments, we adopt a simple decision task where we ask the subjects to choose from the presented contents, which in our case are unreleased movies. The task question is "If you are given a ticket, which movie would you want to see?".

4.1 Contents

4.1.1 Dispersion

The system displays four contents on the screen divided into four equivalent areas. Each area can be controlled independently. Each content consists of five pages of articles. We need to design four categories of contents that are most likely to separate the interest of subjects. After several preliminary trials, we discovered that the most important parameter is the proximity between these categories. Indeed, for the four categories which share nothing in common, the subjects require more intense reasoning to evaluate differences, since there are few correspondences between them. On the other side, when the four categories are really close and present limited dispersion, the subjects cannot evaluate the differences unless they have a very thorough knowledge of the field. For these reasons, we adopt one genre of movies per category. That is, the system presents four different genres of movies. As people usually make a choice between various different genres, the same can be applied to movies. To sum up, each content represents a different movie in genre from the others.

4.1.2 Nature

All the contents that are presented to the subjects are unknown movies, which have not been released yet when the experiments are conducted, to ensure the fairness in the four areas. Each article is of fixed size, i.e., 200 pixels in width and 300 pixels in height, and is in 24-bit color, which is made up of a picture (upper part) and text (lower part). The picture is a production still from the movie, whose area has 180 pixels in width and 100 pixels in height or 120 pixels in width and 150 pixels in height. The text description is written in Japanese for the Japanese subjects or in English for the non-Japanese subjects. The Japanese text is made up of about 100 characters so that it takes approximatively the same time to read each article, and the English text is likewise made up of 6 or 7 lines. An example of the four movies is shown in Fig. 4.

The picture information is representative of the genre as much as possible. Action movie pictures contain fights or cars, whereas comedy movie pictures show smiles or goofy characters. The text information gives some clues about the storyline, the cast, and so on, which is written in the same style as movie reviews in specialized magazines.

We should not make the articles too eye-catching because exogenous attention is very sensitive to sudden changes in luminance. However, it is difficult to collect pictures that do not appear eye-catching. We hence adjust the background color of the whole screen to dark gray and design the white background area of the article to be larger

[†]Height of 1106 mm and width of 622 mm, XGA.

[††]Point Grey Research IEEE1394 camera Grasshopper: UXGA, 30 fps, 8-bit gray scale.

Fig. 4 Example articles of four movies (refer to Internet Movie Database, http://www.imdb.com).

than the picture area, in order to increase the visual fairness between article areas.

As we are using a large screen, the subjects might miss some events when the system updates content located diagonally opposed to the content they are gazing at. We need to always make them aware of the events without aggressively triggering their attention. This is accomplished by framing the last updated article with a black border.

4.2 Time Line

The first phase, i.e. the glimpse phase, is executed to achieve two purposes: letting the subject build the cognitive map and arousing his/her interest in the contents. All articles are displayed at 4-sec intervals in turns (3-sec intervals for the English text). We set the intervals according to the amount of information contained in an article. In the second phase, i.e. the observation phase, the articles are redisplayed in random order at 10-sec intervals (8-sec intervals for the English text) to give the user sufficient time to read them. The time line for the experiment is given in Fig. 5.

1. In the introduction phase, the system displays five frames to explain what it will do for the subject, what situation they are in, what it will ask the subject to do as a task, and how it will display the contents. Each of these presentation events occurs at the label *task*.
2. The label *title* denotes an event for displaying basic information (title, genre, and poster of a movie) in four areas.
3. In the glimpse phase, the respective five articles on four movies (a, b, c, and d) are displayed in one area after another, from the top left to the bottom right ("*TL*" stands for top left and "*BR*" for bottom right, etc.).

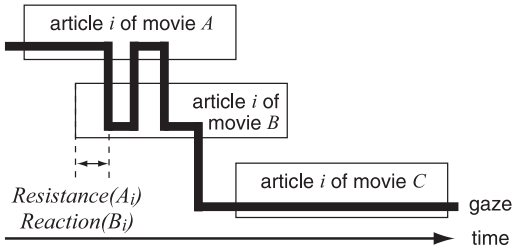

Fig. 6 Timing structures between content presentation events and eye movements. There is an example of two dynamic indicators during the article redisplays of movies *A*, *B*, and *C*.

4. In the observation phase, labels *rand* mean that articles are redisplayed in random order as the order of updates for the four movies has well-balanced context. The order of updates for the five articles of a movie is the same as that in the glimpse phase. Each article remains in the area until the next update occurs in the same area.

4.3 Measurements

We define two dynamic indicators only during the observation phase to measure the responses of subject's gaze to the article redisplays. They are in Fig. 6.

- *Reaction*(x) : The response time to switch the gaze from the previously gazed article to the next redisplayed article x.
- *Resistance*(x) : The delay interval of the gaze fixing on the previously presented article x regardless of the next presentation event. This has the same value as *reaction*, but is credited to the previously gazed article x.

In Fig. 6, where the subject is shifting his/her gaze to the area of movie "*C*" before its article is updated, *Reaction*(C_i) and *Resistance*(B_i) have negative values.

Also, we define *duration* and *frequency* as two baseline indicators for estimating the interest.

- *Duration*(x) : The elapsed time of the gaze fixing on article x.
- *Frequency*(x) : The number of gazes shifting to article x.

In Fig. 6, the subject has gazed at article i of movie "*A*" twice. Therefore, *Frequency*(A_i) is 2 and *Duration*(A_i) is the total duration for the two gazes.

To validate our hypothesis, we correlate these indicators with subjective evaluations (graded with 5 rankings, i.e., 1: no interest, 2: not much interest, 3: indifference, 4: interest, 5: a great deal of interest) through the survey interviews after each task.

5. Experimental Results

5.1 Data Analysis

We conducted four experiments on each of 10 subjects (5

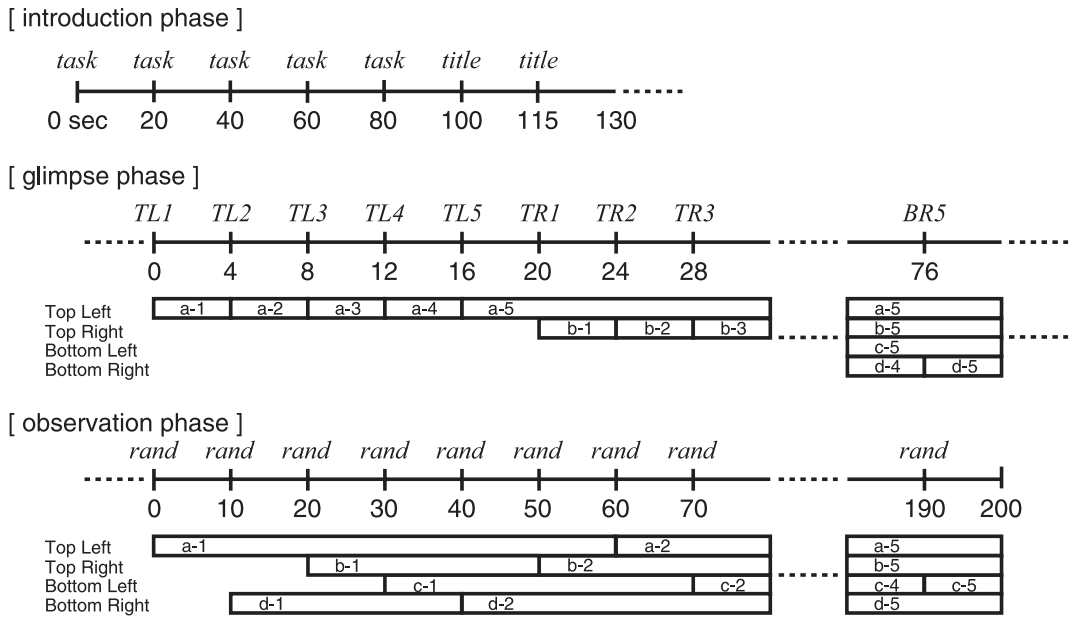

Fig. 5 Time line for the display of the respective five articles on each of the four movies (a, b, c, and d). Labels *task* denote events for displaying task instructions and *title* denote events for displaying basic information about the four movies. "*TL*" denotes that an article is presented in the top left area, "*TR*" in the top right, and "*BR*" in the bottom right. Labels *rand* mean that articles are redisplayed in random order. Rectangular bars, which include symbols "a-♯", "b-♯", "c-♯", or "d-♯" represent the duration of display of articles. Also, symbol "a-1", for example, means the first article of movie "a".

Japanese and 5 non-Japanese). We prepared four movies whose genres were action, animation, comedy, and drama for each of the experiments, i.e., 16 movies for each subject. Each genre was displayed in a different area for each experiment, so that the positional relationship between the genres did not have a bias. Each of the four movies was represented by five articles. Hence, each experiment featured 20 article redisplays (4 movies * 5 articles). We computed the average of *reaction*, the average of *resistance*, the total of *duration*, and the total of *frequency* per movie for the individual subject. We also evaluated whether these indicators were related to the movie selected by the subject and whether they were related to some movies with a higher ranking in the subjective evaluation.

5.1.1 Accuracies of Interest Estimation Using Two Dynamic and Two Basic Indicators

Table 1 lists the accuracies for estimating interest that were calculated as follows: we counted 1 point for the indicator *reaction* if a movie with the shortest average *reaction* was the most interesting that the subject had awarded the highest score and selected in the survey he/she was interviewed in. Otherwise, we counted 0 point. The counted points were then divided by the number of experiments to yield a success rate for matching. The rate is an estimation accuracy of interest. We also applied the procedure to the other indicators (we counted 1 point for *resistance*, *duration*, and *frequency* in the case where a movie with the longest average *resistance*, the longest total *duration*, and the largest total *frequency* corresponded to the most interesting movie).

Table 1 Accuracies for estimating interest for four indicators. The accuracies are rates at which the movie with the best value for each indicator was the most interesting that the subject had selected. The chance level is 25.0%.

Reaction	Resistance	Duration	Frequency
42.5%	55.0%	35.0%	20.0%

Table 2 Matching rates at which the movie with the best value for each indicator was one of the movies with the highest score in the subjective evaluation.

Reaction	Resistance	Duration	Frequency
50.0%	70.0%	50.0%	32.5%

Table 3 Accuracies for estimating two movies of interest. The accuracies are rates at which both movies with the top two values for each indicator were ranked in the top two in the subjective evaluation. The chance level is 16.7%.

Reaction	Resistance	Duration	Frequency
22.5%	45.0%	40.0%	20.0%

The table reveals that *reaction* and *resistance* are better indicators than the two baseline indicators. This result supports our hypothesis. *Resistance*, especially, was more often associated with interest than the other indicators. This demonstrated that if the subject was interested in updated content, he/she often fixed his/her gaze on it without regarding the next update of the other contents. However, we could not obtain satisfactory accuracies for all indicators.

We do not only focus on matching the best values for the dynamic indicators with the selected movie because

Table 4 Average *reactions* (msec) of the most interesting movies ("Selected") and the other three movies ("Others"). The values for the other three movies are in ascending order.

	Subject A		Subject B		Subject C		Subject D		Subject E	
	Selected	Others	Selected	Others	Selected	Others	Selected	Others	Selected	Others
Experiment1	2131	2891, 4388, 4434	1213	1097, 1700, 1803	488	522, 532, 666	2400	1893, 2069, 2609	1881	800, 1209, 1747
Experiment2	1912	1475, 2088, 2247	635	738, 1232, 2078	425	59, 478, 566	853	206, 1809, 2143	409	528, 556, 675
Experiment3	3128	3753, 3809, 4538	591	666, 897, 1569	694	331, 434, 741	1681	684, 1334, 2669	692	375, 992, 1012
Experiment4	1949	3537, 3628, 4396	577	684, 815, 1190	634	440, 522, 528	1603	616, 1090, 2553	459	662, 753, 1185

(a) Japanese subjects

	Subject F		Subject G		Subject H		Subject I		Subject J	
	Selected	Others	Selected	Others	Selected	Others	Selected	Others	Selected	Others
Experiment1	1172	778, 1641, 2097	1122	1012, 1572, 1581	718	375, 437, 456	1122	2034, 2293, 4940	1444	872, 916, 1241
Experiment2	1588	744, 1109, 1869	4203	1362, 3110, 4222	459	509, 594, 853	3518	2478, 3606, 4015	956	585, 769, 1044
Experiment3	850	806, 997, 1169	-143	553, 885, 2053	396	1365, 1900, 2190	3247	2428, 3425, 3797	1647	878, 1019, 1125
Experiment4	506	606, 1100, 1359	519	1319, 2425, 3622	1028	412, 450, 606	3556	4544, 4657, 5766	-53	615, 728, 1106

(b) Non-Japanese subjects

Table 5 Average *resistances* (msec) of the most interesting movies ("Selected") and the other three movies ("Others"). The values for the other three movies are in descending order.

	Subject A		Subject B		Subject C		Subject D		Subject E	
	Selected	Others	Selected	Others	Selected	Others	Selected	Others	Selected	Others
Experiment1	2297	4331, 3572, 3528	937	2125, 1578, 1025	791	478, 458, 113	2972	2769, 2006, 1715	1987	1884, 887, 238
Experiment2	2285	2047, 1919, 1281	1138	1773, 1191, 797	641	447, 443, 431	2031	1406, 1334, 1134	743	581, 535, 340
Experiment3	3478	5813, 3199, 1485	821	879, 482, -574	544	478, 297, 141	1840	1553, 1375, 1102	819	632, 563, 559
Experiment4	3684	3928, 3293, 2503	793	1178, 693, 507	581	730, 440, 434	994	2156, 1375, 1328	-28	741, 512, 461

(a) Japanese subjects

	Subject F		Subject G		Subject H		Subject I		Subject J	
	Selected	Others	Selected	Others	Selected	Others	Selected	Others	Selected	Others
Experiment1	2473	753, 225, -1081	1797	1472, 1321, 1053	334	636, 562, 424	2898	2680, 2328, 2222	1664	1490, 647, 459
Experiment2	746	1888, 450, -460	-246	4050, 1177, 856	-653	777, 444, 438	1190	3525, 2000, 1840	919	866, 684, 638
Experiment3	1556	871, -250, -532	1785	1197, 488, -12	590	2821, 640, 12	3587	3122, 2836, 2812	537	1631, 1584, 821
Experiment4	1699	1531, 1146, -1047	5360	4085, 1610, 973	350	853, 659, 543	6338	4071, 3848, 3575	1222	978, -456, -947

(b) Non-Japanese subjects

users are not always interested in only one movie. In fact, some subjects awarded two movies the highest score as reported in the survey. Therefore, we counted 1 point if the movie with the best value for each indicator was one of the movies with the highest score. The matching rates based on this rule are listed in Table 2. We can estimate content with a high degree of interest with 70% accuracy by adopting *resistance*.

There is more than one method of estimating interest using the dynamic indicators. Depending on the strategy of the information service, the system should estimate plural contents of interest. We consider, for example, that the system will estimate two types of contents of interest and select the next information based on these to provide to the user. Table 3 lists the estimation accuracies of two movies of interest. The accuracies are rates at which both movies with the top two values of each indicator were ranked in the top two in the subjective evaluation. *Resistance* was better than the other indicators as with the results for the most interesting movie. However, its accuracy was not good (45.0%). It was difficult to estimate both the top two interesting movies because the chance level was 16.7%. Here, we computed the rate at which the movie selected by the subject was either of the two movies with the top two values of *resistance* and obtained 70.0%. When the system estimates two movies of interest based on *resistance*, we can hit the most interesting movie in a high rate and another movie of interest as often as not.

5.1.2 Analysis of Relation between Two Dynamic Indicators and Interest Based on User's Personality

We consider that the subjects fall into two types of personalities. Tables 4 and 5 list the average *reactions* and *resistances* in the four experiments on the 10 subjects. The "Selected" column has the dynamic indicators for the most interesting movie. The "Others" column has the dynamic indicators for the other three movies.

We focus on seven subjects (Subjects C, D, E, F, G, I, and J) listed in Tables 4 and 5. We can confirm that most of the movies they selected had the longest *resistance*. For the other three subjects, some movies with the shortest *reaction* matched the selected movies. The matching rate for the seven subjects and the other three subjects were 75.0% and 66.7% as listed in Tables 6 and 7. The seven subjects did not tend to shift their gaze to contents of less interest as soon as the contents had been updated. On the other hand, the other three subjects tended to react more quickly to updates of interesting content. If the system can discriminate between the two types of personalities based on some kinds of behavior, we can estimate the user interest with 70% accuracy. In the next step of our work, we intend to focus on whether the two types of user behaviors correlate with other overt and observable user traits, e.g. personal tempo [15].

Table 6 Accuracies for estimating the most interesting movies by subjects C, D, E, F, G, I, and J.

Reaction	Resistance	Duration	Frequency
32.1%	75.0%	42.9%	21.4%

Table 7 Accuracies for estimating the most interesting movies by subjects A, B, and H.

Reaction	Resistance	Duration	Frequency
66.7%	8.3%	16.7%	16.7%

Table 8 Accuracies for estimating disinterest for four indicators. The accuracies are rates at which the movie with the worst value for each indicator was the least interesting.

Reaction	Resistance	Duration	Frequency
32.5%	32.5%	40.0%	40.0%

We consider that the system needs to probe user's personality before *Mind Probing* to estimate interest more accurately.

5.2 Discussion

The dynamic indicators were better than the baseline indicators for estimating interest throughout the experiments. The baseline indicators are more easily influenced by the complexity of contents. Whereas, when a user compares some contents, the baseline indicators reflect interest because he/she frequently gazes at contents of interest. However, the proposed presentation method barely stimulates the user to compare them by switching his/her gaze because it frequently updates some contents. The method induces the user to proactively acquire the necessary information by resisting exogenous stimuli, considering the accuracy was better with *resistance*. Under dynamic-information services, we consider that time delays in interaction between the system and the user would be greatly associated with more interest.

However, we could not estimate the most interesting movie satisfactorily with only a dynamic indicator for all subjects. As described in 5.1.2, the subjects' personalities caused this result. Moreover, variations in the time of interest might have influenced the estimates. We could not also strongly correlate the intensity of interest with the dynamic indicators. Although we verified the contraposition of our hypothesis as an additional analysis to find out the correlation, i.e., whether the least interesting movie corresponded to the longest *reaction* or the shortest *resistance*, this was not true (about 30% success as shown in Table 8). We guess that single dimensional features have limitations. The intensity of interest might therefore be shown on multidimensional space that includes the other indicators.

6. Conclusion

We proposed a novel approach to estimating user interest in dynamic-information services, which was based on the relationship between the dynamics of proactive content presentation and eye movements. According to our evaluation of the model of proactive behavior, we concluded that:

- A user's interest can be estimated via the dynamics of the eye's response to proactive content updates by managing the presentation with two specially designed phases.
- The *resistance* indicator, i.e., the delay interval of the gaze fixing on content without regarding other presentation events, can efficiently be used to estimate the interest of many users.

Our design for presentation that frequently updates contents is very busy, whereas it is better for estimating interest. It may not be comfortable for some users. We need to evolve the design while maintaining a balance between usability and accuracy of interest estimation.

The next step after estimating interest is undoubtedly to use it to provide responses that can be adapted by interactive system. For example, if the system detects that a movie has attracted the user interest strongly more than with other movies, it can provide the user with more detailed information on the movie or recommend other movies related to the interest. The system can also probe user interest more accurately while providing the information that follows. We need to discuss how the information can be chosen to effectively probe the interest. We intend to evolve *Mind Probing*, and build a system that can skillfully provide sensible information to user and increase its degree of integration with humans.

Acknowledgements

This work is in part supported by Grant-in-Aid for Scientific Research of the Ministry of Education, Culture, Sports, Science and Technology of Japan under the contract of 18049046.

References

[1] P.J. Silvia, Exploring the psychology of interest, Oxford University Press, 2006.
[2] W. McDougall, An outline of psychology, Sigaud Press, 2007.
[3] M.A. Just and P.A. Carpenter, "Eye fixations and cognitive processes," Cognitive Psychology, vol.8, pp.441–480, 1976.
[4] L.E. Sibert and R.J.K. Jacob, "Evaluation of eye gaze interaction," Proc. SIGCHI Conference on Human-Factors in Computing Systems, pp.281–288, 2000.
[5] H. Deubel and W. Schneider, "Saccade target selection and object recognition: Evidence for a common attentional mechanism," Vision Research, vol.36, no.12, pp.1827–1837, 1996.
[6] M. Posner, "Orienting of attention," Quarterly Journal of Experimental Psychology, vol.32, no.1, pp.3–25, 1980.
[7] R.W. Picard, E. Vyzas, and J. Healey, "Toward machine emotional intelligence: Analysis of affective physiological state," IEEE Trans. Pattern Anal. Mach. Intell., vol.23, no.10, pp.1175–1191, 2001.
[8] A. Kapoor, R.W. Picard, and Y. Ivanov, "Probabilistic combination of multiple modalities to detect interest," Proc. 17th International Conference on Pattern Recognition, vol.3, pp.969–972, 2004.

[9] P. Qvarfordt and S. Zhai, "Conversing with the user based on eye-gaze patterns," Proc. SIGCHI Conference on Human-Factors in Computing Systems, pp.221–230, 2005.
[10] T. Onishi, T. Hirayama, and T. Matsuyama, "What does the face-turning action imply in consensus building communication?," Proc. 5th International Workshop on Machine Learning for Multimodal Interaction, pp.26–37, 2008.
[11] E.C. Tolman, "Cognitive maps in rats and men," Psychological Review, vol.55, no.4, pp.189–208, 1948.
[12] R.M. Downs and D. Stea, "Cognitive maps and spatial behavior: Process and products," Image and Environment, pp.8–26, 1973.
[13] T.F. Cootes, G.J. Edwards, and C.J. Taylor, "Active appearance models," Proc. 5th European Conference on Computer Vision, vol.2, pp.484–498, 1998.
[14] B. Triggs, P. McLauchlan, R. Hartley, and A. Fitzgibbon, "Bundle adjustment — A modern synthesis," Vision Algorithm: Theory & Practice, B. Triggs, A. Zisserman, and R. Szeliski, eds., Springer-Verlag LNCS 1883, 2000.
[15] H. Rimoldi, "Personal tempo," J. Abnormal and Social Psychology, vol.46, pp.283–303, 1951.

Takatsugu Hirayama received his B.Eng. in electrical and information engineering from Kanazawa University, Japan, in 2000, and his M.Eng. and D.Eng. in engineering science from Osaka University, Japan, in 2002 and 2005. He is currently a visiting researcher in the Graduate School of Informatics, Kyoto University. His research interests include facial image recognition, human communication, and human-machine interaction. He is a member of the Information Processing Society of Japan and the Human Interface Society of Japan.

Jean-Baptiste Dodane received his M.S. in engineering from Ecole Centrale de Lyon, France, and his M.S. in informatics from Kyoto University, Japan, in 2009. His research interests include human communication and human-machine interaction.

Hiroaki Kawashima received his M.S. and Ph.D. in informatics from Kyoto University, Japan in 2001 and 2007. He is currently a lecturer in the Graduate School of Informatics, Kyoto University. His research interests include time-varying pattern recognition, human-machine interaction, and hybrid dynamical systems. He is a member of the Information Processing Society of Japan, the Human Interface Society of Japan, and the IEEE Computer Society.

Takashi Matsuyama received his B.Eng., M.Eng., and D.Eng. in electrical engineering from Kyoto University, Japan, in 1974, 1976, and 1980. He is currently a professor in the Graduate School of Informatics, Kyoto University. His research interests include knowledge-based image understanding, computer vision, cooperative distributed vision, 3D video, and human-machine interaction. He has received nine best paper awards from Japanese and international academic societies including the Marr Prize at the International Conference on Computer Vision in 1995. He is a fellow of the International Association for Pattern Recognition and the Information Processing Society of Japan, and a member of the Japanese Society for Artificial Intelligence and the IEEE Computer Society.