PsychNology Journal, 2007
Volume 5, Number 2, pp. 133 -164

Development and Evaluation of a Method Employed to Identify Internal State Utilizing Eye Movement Data

Noriyuki Aoyama ♠♦ and Tadahiko Fukuda ♣

♦ Graduate School of Media and Governance, Keio University (JAPAN)

♣ Faculty of Environmental Information, Keio University (JAPAN)

ABSTRACT

In the attempt to recognize and estimate human internal states, such as varying emotions, psychological and conceptual conflicts pose interesting and challenging issues. In this paper, we explore a pattern recognition technique that can detect a state of confusion and can estimate human interest, each an internal state of mind. In order to automatically detect a state of confusion from the objective data made available to us, the technique we present relies upon eye movement data. We have conducted three experiments in which subjects are confronted with a task that includes a trap intentionally designed to confuse them. We have recorded their eye movement data. We demonstrate that approximately 89% of a state of confusion can be detected from eye movement data by using a backpropagation algorithm. Moreover, for estimating human interest, we present a technique that builds upon the foundation of our confusion detection technique. As a result, we can demonstrate that approximately 60% of human interest can also be estimated through eye movement data.

Keywords: *Confusion, Interest, Eye Movement, Human-Computer Interaction, Neural Networks.*
Paper Received 26/07/2006; received in revised form 22/05/2007; accepted 03/08/2007.

1. Introduction

As computer systems continue to improve upon their already high levels of performance and sophistication, the corresponding interactive systems that process information based upon user input are being utilized for numerous applications that improve functionality. This improvement process is predicated upon the fact that users tend to view sophisticated interactive systems as a servant who serves to satisfy various user demands (Reeves & Nass, 1996). Although originally intended to improve

♠ Corresponding Author:
Noriyuki Aoyama,
Graduate School of Media and Governnace, Keio University,
k205, 5322 Endoh, Fujisawa-shi, Kanagawa, Japan
E-mail: aonori@sfc.keio.ac.jp

interactive system functions for the better, ironically, these improvements have also increased the chances of creating user experiences in which complex problems are encountered, ones that cannot be solved by the users themselves. Most users expect the functionality of an interactive system to be flexible, yet scenario–specific, and responsive to any given situation. This very concept sets up the pitfall users can stumble into during the process of attempting to accomplish their program goals. When users are faced with interactive systems, they will attempt from the outset to work out quick strategies to accomplish their goals. The users then proceed in a methodical manner in pursuit of their goals. This is a perfectly acceptable strategy when the process progresses favourably and satisfactorily. However, if the interactive system generates an output that is different from a user's expectation, the user becomes confused and can even suffer from psychological stress. As a consequence of this experience, some users may go so far as to try to avoid the use of interactive systems altogether. In a confirmed trend, some users have been found to feel the same type of responsibility that a manufacturing company may feel when a functionality or safety problem occurs, as well as feeling a similar responsibility when usability and convenience problems occur (Norman, 1988). Whether or not an individual will experience this phenomenon is dependent upon his/her unique sense or consciousness. That is, users can experience conflicts between their desires with respect to convenience improvements and their desire to use the right products and their overinflated confidence about their literacy in utilizing such convenient products. Indeed, these factors often stand in the way of completely successful user utilization of an interactive product. This dynamic is also made clear when examining the choices that industrial manufacturers make in adopting more human-centred product designs.

Firstly, the concept of human-centered design was originally proposed by Norman and Draper (Norman & Draper, 1986) and it became the front line of product development, preferencing utility over usability. This concept has since been propagating the model of giving more weight to actual usage scenes within the product design process. This user-oriented concept has now become a fundamental product development concept.

Secondly, the human-centred design concept has favourably enhanced user satisfaction in terms of the notion now referred to as "usability". It was proposed by Nielsen (Nielsen, 1993), who had been inspired by several other human-centred design related concepts, such as universal design (Mace, 1985), cognitive engineering (Card, Moran, & Newell, 1983), and Kansei engineering (Nagamachi, Ito, & Tsuji,

1988). The usability concept, which incorporates three other fundamental concepts, is absolutely essential for product development.

Finally, manufacturers must currently attend to usability-satisfied users' increscent demand for interactive systems. That is, the market demands ever more personalized designs befitting individual users and their unique situations. The term "adaptive" best describes this new concept. Since the late 1990s, the term "adaptive user interface" has been used to describe this new interaction design within human-computer interactions (HCI) (Maglio, Barrett, Campbell, & Selker, 2000).

To better realize this adaptive user-interface, many researchers from a number of fields, such as artificial intelligence, ergonomics, and HCI, are working to recognize and estimate people's internal landscapes. These researchers are pursuing the possibility of developing diverse new interactive systems. For example, several research groups have attempted to build robots that possess human-like emotions (Hara & Kobayashi, 1996; Lim, Ishii, & Takanishi, 2000; Zecca et al., 2004). Conversely, another researcher is working to develop intelligent systems (e.g., an automatic breaking system for automobiles) that are designed to avoid dangerous or hazardous situations originally caused by human error. The critical factor is that these adaptive interactions need to operate automatically without the intervention of any user autonomous behaviours and intentions. For instance, when a user uses an interactive system (e.g., a Japanese bank ATM, which is highly complex in comparison to ATMs in other countries), that system should, first of all, provide the shortest possible path to the user's goal and preclude any information that is not indispensable to that user at that moment. If the user is unable to attain his/her preset goal as a result of problems, the system should automatically present new, detailed, and useful information that can reverse the situation and once again place the user on the right path to his/her goal. Consequently, the amount of time needed for a user to achieve the goal will overall decrease.

In this study, an attempt was made to develop a method that can identify user internal states by utilizing eye movement data during HCI. In order to conduct this research, the term *internal state* had to first be defined. Most relevant research studies in this field that had aimed to identify the "internal state" of humans had focused on identifying emotions that many researchers catalogued as basic emotions. However, basic emotions and research issues related to them had not been specifically defined, although researchers have been attempting to define emotions since the days of Plato and Aristotle dating back to circa 400 B.C. This research issue has been addressed

contemporarily by Tomkins, Ekman, Izard, and others (Tomkins, 1962, 1963; Izard, 1991; Ekman, 1992). Although one might say that this type of work has been ongoing for more than two millennia, even the most basic emotions, serving as building blocks for the more complex emotions, remain undefined. The reason for this lies with the lack of clear and numerical criteria for using these terms that indicate an emotional state. In addition, emotional states differ in meaning according to who is experiencing them and in what context. Therefore, in this study, a specific definition of *internal state* was avoided. On that basis, *internal state*, in this study, will not be defined in detail but will instead be defined in the broadest of terms. The term will be defined in terms of a specific situation and context, such as the happiness of accidentally encountering best friends. *Internal state* eventually became defined in this work as those physiological and psychological state changes caused by the acquisition of external information and the internal processing of that information. In particular, the internal states of confusion and interest were focal points for us within the framework of potential internal states, because both confusion and interest are practical and highly versatile internal states when considering the real world of HCI.

Our hypothesis in this study was that the changes in a person's internal state are somehow reflected in his/her eye movement data, and this is the reason why we employed eye movement data to detect people's internal states in this study. Our hypothesis was proposed based upon our consideration of three factors: (1) previous studies conducted in related research topics; (2) our modification of the research concept for our study; and (3) the ease of measurement for our study.

Previous research has determined, as acknowledged in the first consideration point for our hypothesis, above, that although internal state information is not contained within eye movement per se, but that the eye itself, e.g. pupil size, indicate the changing human psychological makeup. Hess conducted experiments indicating that pupil size changed in accordance with the subject's interest in objects being viewed (Hess & Polt, 1960; Hess, 1965). An interpretation of this result can be that human interest affects their eye pupil sizes and this interpretation pointed to our own hypothesis. However, there was no previous study in which eye movement data was directly connected to changes within individual internal states.

Our second point in considering our hypothesis was the modified research concept. The traditional research concept utilizing eye movement data focuses more on how humans acquire visual information. The traditional concept focused on the activity of human information acquisition (e.g., Byrne, Anderson, Douglass, & Matessa, 1999;

Hayhoe & Ballard, 2005). In contrast, our new concept is based upon how humans communicate their internal information, bringing it into view through their eyes. In general, humans communicate more easily with other people in face-to-face situations (Baron-Cohen, Wheelwright, & Jolliffe, 1997; Driver et al., 1999; Macrae, Hood, Miline, Rowe, & Mason 2002). This concept of face-to-face communication focuses on the activity of human information-providing. In addition, we found biological support for our hypothesis; for instance, human eyes have unique characteristics, such as the proportion between the exposed white sclera and the surrounding darker-coloured iris, the largest width-height ratio within the eye's outline, and the whitest of exposed sclera among all primates, which enables communicating with ease by using a gaze signal (e.g., Morris, 1985; Kobayashi & Kohshima, 1997; Kobayashi & Kohshima, 2001). Finally, the reason for using eye movement data is that such data can be easily collected from subjects without direct physical contact. The aforementioned factors guided us in the formation of our hypothesis.

In previous studies, three primary methods have been used to identify internal states by recording external data without any need for direct physical contact. The first method recorded eye movement related data; the second method recorded facial expression data, and the third recorded speech data. Umemuro and Yamashita (2003) attempted to detect states of confusion and states of surprise by assessing pupil diameter data within HCI. These authors defined users as being in a state of confusion whenever any unexpected disturbance obstructed the users from continuing their task and those users were unable to find an appropriate solution for handling that situation. Surprise, in contrast, was defined as the status of users upon recognizing an unexpected event, but such users were still able to continue their task. Researchers collected pupil diameter data from memory task trials, and then analyzed the data distribution. As a result, the subjects' confusion and surprise could be detected at rates of 25% and 65%, respectively. However, the pupil diameter data acquired with this method were inadequate because such data included the interference effects of the surrounding environment and other visual stimulation. For example, the amount of ambient light, as well as the colour and brightness of visual objects, all affect and impede pupil diameter data.

There have also been studies utilizing facial expression data (Yacoob & Davis, 1996; Lyons, Akamatsu, Kamachi, & Gyoba, 1998) and speech data (Dellaert, Polzin, & Waibel, 1996; De Silva & Ng, 2000; Nicolson, Takahashi, & Nakatsu, 2000). Yacoob and Davis explored facial expression recognition by using optical flow data. They

focused on long-term image sequencing to recognize seven facial expressions—six different facial emotions and one “neutral” face—by integrating spatial and temporal information from video images. However, the recognition rate for the emotions of fear and sadness was low in comparison to the relatively high recognition rates for the emotions of happiness, anger, and surprise. Lyons et al. (1998), who conducted another study using facial expression data, tried to extract and code facial expression information from images by using a Gabor filter and calculating the geometric aspects of facial expressions. Lyons et al. analyzed the correlations between the Gabor-coded data and the semantic or geometrical ratings assigned to the respective stimulant images of the subjects participating in the experiment. They found the average correlation rate of a Gabor model to be approximately 0.57 and that of a geometrical model to be approximately 0.37.

These facial expression studies illustrate the same problem that affects the current recognition approaches being tested. These methods cannot recognize natural expressions. They are only capable of recognizing those artificially reproduced expressions that subjects have been asked to make with specific facial movements to display expressions for various emotions.

Using speech analysis, Dellaert et al. (1996) explored statistical patterns in an attempt to recognize four emotions based upon the pitch data of test subjects’ speech. Dellaert et al. analyzed the data by comparing the performance of their proposed recognition technique in relation to human performance. The result was that the recognition technique indeed had some success with the carefully orchestrated human test performances. However, this technique had substantially high error rates that ranged from 20% to 44%. Nicolson et al. (2000) tried to recognize eight emotions by analyzing the speech data within HCI. They used neural networks to recognize emotions using both phoneme and prosodic features data. As a result of the analysis of the recognition rates of three networks namely, One-Class-in-One, All-Class-in-One, and Learning Vector Quantisation, the highest recognition rate recorded was approximately 57% for the All-Class-in-One neural network. This study served quite favourably as a reference but, nevertheless, the method was not sufficiently well adapted to use HCI without speech. Moreover, De Silva and Ng (2000) tried to recognize six basic emotions using both the data of audio waveforms to analyze emotional speech by Hidden Markov models and the data of video images to analyze facial expressions using statistical techniques. The recognition rate extracted from video-image data was approximately 65%, and the rate for audio data was

approximately 35%, with a 72% rate for bimodal data. Although those recognition rates were relatively high, it is still difficult to say that this system is practical because it requires creating a user-by-user recognition model.

Studies that require physical contact for acquisition of the data necessary for estimating feelings and recognizing emotions have also been conducted (Musha, Terasaki, Haque, & Ivanitsky, 1997; Ishino & Hagiwara, 2003; Takahashi, 2004). Ishino and Hagiwara tried to estimate four different emotional states by reading electroencephalography (EEG) data and using a neural network. Their estimation rate was approximately 61%. Takahashi tried to recognize the states of pleasure and displeasure from EEG data by using a neural network and a support vector machine. The higher recognition rate was 62.3% for a neural network classifier. These results indicated that EEG could certainly provide a cluster of useful data for understanding internal states. However, the current devices used to acquire EEG data are impractical and unrealistic for use in HCI applications because such devices are prohibitively clumsy and disruptive. In addition, subject internal states could have been adversely affected by the extensive and complicated nature of the measurement system method itself.

Eye movement data can be provided without any physical contact using the above-described equipment and methods that were utilized in previous studies. However, because most commercially available eye tracking systems presently offered use infrared rays to detect eyes' positioning and orientation, it is necessary to ensure a lighting environment that is measured with absolute precision. A new technology that can adjust to a changing light environment has been recently developed using image processing technologies and algorithms (Yamanobe, Taira, Morizono, & Kamio, 1990; Zhu, Fujimura, & Ji, 2002). These features provide great advantage in working with internal states using eye movement data. Alternatively, the eye tracking method also poses a disadvantage. Namely, all existing equipment used to record eye movement requires calibration before commencing any eye movement recording. This calibration is necessary in order to measure eye movement with the potentially highest accuracy. This is a technical issue that many researchers have attempted to address (e.g., Ohno, 2006). However, in the case of a contact-free eye tracking system, the calibration procedure is easier for the researcher and less burdensome for the subjects when compared with other equipment and methods that could also be used to acquire objective data. In fact, eye movement data can be considered to be the most reliable and comprehensive method for assessing the internal state of a human being.

If another method could be developed that identifies human internal states using data obtained without any physical contact, then those systems that users interact with would be able to provide appropriate information to fit the use context and situation depending upon the user's internal state. This would enable not only the possibility of increasing users' chances of troubleshooting problems on their own without needing system help, but it could also decrease the chance of users experiencing psychological stress.

2. Confusion Detection Method

This term "confusion" had to be defined at the outset, before conducting any experiments. It is difficult to define the meaning and express it in a single word because of the high variability of human internal states. This is the very reason why researchers typically use more than one word to describe an internal state and the terms used to describe an internal state have varied by subject. For example, there are certain words that are related to confusion, such as puzzlement, embarrassment, and panic. However, these words have been used appropriately in accordance with each particular situation. Therefore, in this study, the meaning of the word "confusion" has been defined in relation to the actual context in which such confusion has occurred.

2.1. Experiment 1

Our first experiment was conducted to demonstrate the fact that eye movement data can be used successfully to detect a person's internal state. This experiment was designed to recognize a state of confusion from eye movement data in a context without any physical contact, while the interaction of multiple experiences among subjects still was taken into consideration.

2.1.1. Experiment System

The experiment system consisted of two personal computers (PCs). The first PC was used to output stimuli onto its monitor. The other PC controlled the eye tracking system. A simple system configuration is shown in Figure 1. Stimuli were presented on a SyncMaster 172MP-R (Samsung Japan Corporation, Tokyo, Japan) with a 17-inch liquid crystal display. Eye movement data was gathered using the image and the cursor, in which the subjects' sight points were superimposed. To record the

superimposed images, two SSC120EX scan converters (Canopus Co., Ltd., Hyogo, Japan) were used. These images were then recorded by a Sony WV-DR7 (Sony Corp., Tokyo, Japan) DV/VHS combo cassette recorder. The eye movement data obtained from the dominant eyes were measured using the FreeView eye tracking system (Takei Scientific Instruments Co., Ltd., Niigata, Japan). The FreeView is a non-contact eye tracking system, which enabled easier operation of the computer mouse by seating each subject on a high-backed chair with a headrest, where each subject was instructed to place his/her head on the headrest to establish the correct distance between the eye camera and the subject's eye. This experimental setup is illustrated in Figure 2. The distance from the subject's eye was fixed at 70 cm. The FreeView system tracked each subject's eye movement automatically by employing a pupil-corneal reflection method using an infrared ray at a sampling rate of 30 Hz. This number was dependent upon the NTSC frame rate. In this paper, the word "frame" is always used in this context. The precision rate was far less than 0.1 degrees, even under the worst conditions, and the measurable range was ±20 degrees in both horizontal and vertical directions.

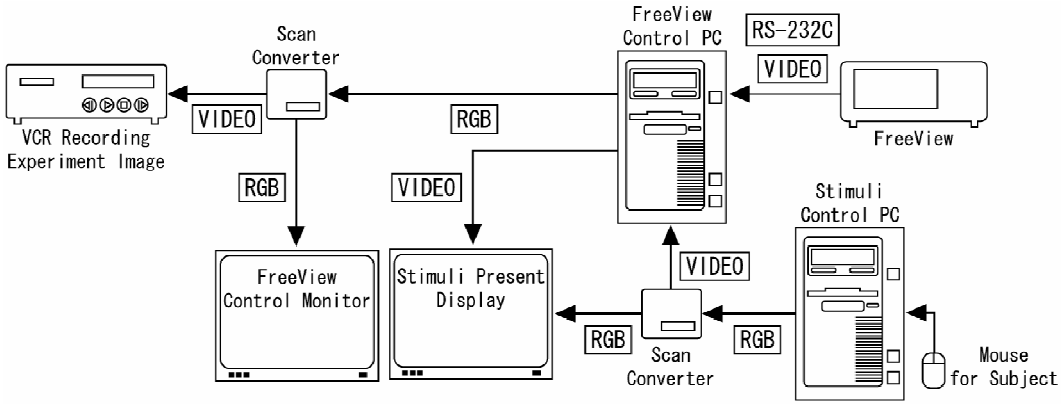

Figure 1. Configuration of the Experiment System. The experiment system consisted of a detection station for measuring eye movement, two PCs with displays for controlling the FreeView system and presenting stimuli, two scan converters for converting signal and branch signals, and a video cassette recorder.

Figure 2. Experimental Setup. The subjects sit in front of the display without any device making physical contact with them. Eye movements are measured while each subject performs tasks by manipulating a computer mouse.

2.1.2. Task and Procedure

In Experiment 1, the types of confusion caused by two different scenarios were identified as described below:

(1) The confusion caused by interrupting a plan that the subject has simulated; and

(2) The confusion caused by displaying a result differing from what each subject has anticipated.

Four types of stimuli, labelled as Task 1 through Task 4, were constructed for use with simple computational problems. Task 1 and Task 2 were designed to stimulate the first type of confusion described above, and Task 3 and Task 4 were designed to stimulate the second type of confusion. Each stimulus included a trap designed to confuse the subjects. The type and details of each stimulus are shown in Table 1. Sample screens of the first and last screens (the trap screen) in Task 2 are shown in Figure 3.

Task	Type	Detail
1	Adding and/or subtracting numbers to achieve a target value	The screen displays a present value and a target value, as well as two buttons, one assigned with a positive number and the other with a negative number. The subjects will try to reach the target value by selecting one button. They need to select the positive number because the present value is lower than the target value. However, in the process, subjects will notice that the positive number button option, when selected, creates a result exceeding the target value.
2	Answer selection of computational problems	The screen displays several computational problems and subjects must choose from among three answers. The next problem will be presented only when the subjects have correctly answered the current problem. During this process, however, the subjects will realize that no correct choice is being offered to solve the problem at hand.
3	Answer selection of computational problems	The screen displays several computational problems and subjects must select from among three answers for each problem. The next problem will be presented only when the subjects have correctly answered the current problem. During this process, however, the subjects will face a problem in which all the choices offered are the correct answer.
4	Adding and/or subtracting values to achieve a target value	The screen displays a present value and a target value, as well as two buttons, one assigned with a positive number and the other with a negative number. The subjects will try to reach the target value by selecting one button. They need to select a positive number button because the present value is lower than the target value. However, during this process, the subjects will discover that the selection buttons do not have any number at all assigned to them.

Table 1. Type and Details of Each Stimulus in Experiment 1

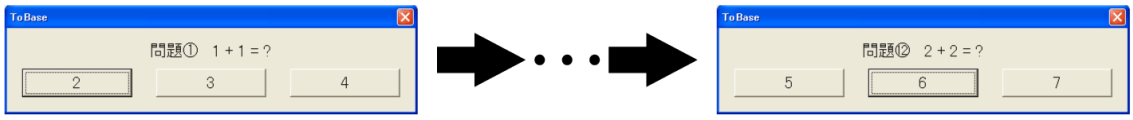

Figure 3. Sample Screen Shot of Task 2. The subjects were asked to choose the correct answer for the calculation problems (left). However, no correct answer is among the choices for the problem at hand (right).

The subjects of this experiment consisted of 65 males and 31 females. They were both undergraduate and graduate students between the ages of 18 and 24 years old. All reported that they had normal vision, including corrected eyesight.

Each subject carried out only one randomly selected task. And, in pilot test, several subjects were asked to comment few words about the experimental procedure and the tasks after the experiment for any revisions that may need to be made for the actual experiment. In doing so, the issues of interaction between tasks and the possible psychological impact of the task including a trap to stimulate the eye movement data were all considered. The breakdown of the final number of subjects assigned to each is shown in Table 2. In this Table, the subject number of Tasks 2 and 4 was larger than Tasks 1 and 3 because the large number of subjects was assigned to Task 2 and 4 based on the subjects' comment that among the four stimulant tasks, the operations within Tasks 2 and 4 were the two most comprehensible ones as a stimulus. The more comprehensible the task, the acquired data from the experiment was thought to be easier to recognize the pattern. Therefore, this experiment was conducted, based on the classification of two types of simple confusion state above, the variance of each type of simple confusion state and each Task were not considered to have major influence on the result of eye movement output by humans.

However, in the post-pilot test interviews, the subjects commented that among the four stimulant tasks, the operations within Tasks 2 and 4 were most easily comprehensible as stimuli. Therefore, in order to achieve our objectives and to accomplish our mission, Tasks 2 and 4 were given more emphasis within this experiment's structure. The breakdown of the final number of subjects assigned to each task is shown in Table 2.

Task	1	2	3	4
Number of Subjects	11	45	11	29

Table 2. Breakdown of the Number of Subjects Assigned Each Task

2.1.3. Detection Method

Based upon our hypothesis that internal state changes are somehow reflected in eye movement data, we included obtaining data about velocity changes for lines of sight within this study. This decision was based upon several sets of analyses acquired as a

result of conducting the pilot test. A velocity change in a line of sight was calculated by using the difference from the previous frame. The data set of each frame was calculated by using the sight point location data shown on the display as acquired through the FreeView system. Degrees-per-second was adopted as the measurement unit because the sight point location data measured the degrees from each subject's eyes as a unit.

In order to create the confusion detection model, we attempted to use a learning neural network that utilized eye movement data as input data for extracting various features of a state of confusion when subjects encountered stimulating traps. Before compiling this report, we employed approaches that focused on the data of each frame as acquired through the FreeView system, but we were unable to grasp the specific difference between normal phase data and trap phase data that can detect a state of confusion. Therefore, the learning algorithm of a neural network was employed to treat eye movement data as a cluster of historical data extended to multiple frames to consolidate the features of a state of confusion during a relatively long duration. The model created could serve as a template and would include detecting the features of confusion as acquired from eye movement data that are considered as the time-series data.

2.1.4 Neural Network Learning Algorithm

The neural network used for learning in this study is a backpropagation network algorithm using the sigmoid function. This network had eight layers, including six hidden layers. The construction of units in each layer is shown in Table 3.

Layer	IL	HL1	HL2	HL3	HL4	HL5	HL6	OL
Number of Units	50	100	50	40	30	20	10	1

Table 3. Construction of Units in Each Layer

The input data mass number was set at 50 frames. This number was based upon the metal operator proposed by Card, Moran and Newell (1980). We set the time-duration for approximately 1.67 seconds. However, the setting of the metal operator was approximately 1.35 seconds. Because our setting scenario manipulated information through use of the Internet, it was a more complex one than the Card et al. setting, which simulated a simple HCI situation with a computer mouse.

The learning policy was such that the data set for the first 50 frames at the time of the subjects' initial encounter with the trap screen would be fixed to "1" as a target output. The other data sets for before the subjects encountered the trap screen would be fixed to "0" as a target output. However, all the data, excluding the first 50 frames of the trap screen phase, were not used in the learning because they would be considered to be noise in the learning of the neural network. The data input was carried out in such a manner as to input the neural networks by delaying one frame at a time. This learning policy is shown in Figure 4. The data was normalized while the input of the data of line of sight was carried out. The learning rate in the learning process was set at 0.001, and the neural network ceased learning when the value of the cost function reached 0.00001.

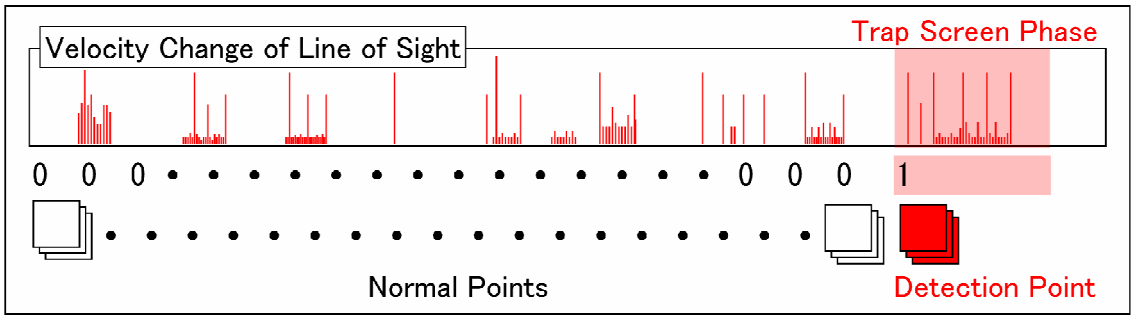

Figure 4. Learning Policy of the Experiment 1 Neural Networks. The unit of one data set was fixed at 50 frames. The data set other than the expected detection point was labelled as "0". For the anticipation detection data set, a label of "1" was given. The data set labelled "1" was counted from the point where the trap screen was presented and continued for 50 frames beyond that point.

Because the experimental objective was to develop a method for detecting the basic features of a state of confusion, the neural network learned by grouping together the four types of data that were acquired from this experiment.

2.1.5 Detection Results

As a result of the neural network learning using approximately 70% of the data acquired during the experiment, the learning coalesced and thus the confusion detection model was created. To confirm the detection accuracy of the model created, verification was conducted using approximately 30% of the data as testing data, data that had not been previously used for the neural network learning. A detection condition was judged to have occurred when the output unit of the confusion detection model provided a positive value at the points where it was expected that a subject's confusion would be detected. Therefore, we focused only on the results detected at

those points where the detection of confusion was anticipated. We decided not to discuss false detection results, such as any output values of less than 0.5 at random points other than those points where it was expected that the output unit would make its detection. Moreover, in this report, the detection rate refers to the detection ratio of the anticipated points from the eye movement data as time-series data created by the model.

Verification showed that the detection rate for testing data was 100%. That is, all points that should have been detected as causing a state of confusion were indeed detected. The sample results are shown in Figure 5. The figure is comprised of three charts: the first chart (top) are the velocity changes within the lines of sight; the second chart (middle) shows the points where a state of confusion was expected; and the third chart (bottom) shows the output unit values. There are numerous peaks in Figure 5 because the verification was conducted in such a manner as to combine all the variables into one chart. However, only one point per subject is anticipated.

In this verification, the control situation with “no confusion” was not treated and discussed, although the reliability could improve if the same pattern with “confusion” were found from the data with “no confusion”. Since the neural network was learned by treating 50 frames as a data unit and delaying one frame at time, and then number of points given “1” as target output of a neural network on eye movement data was extremely smaller than the number of points given “0”, the reversing of our proposed learning algorithm in this study was not conducted and was not deal with in this paper.

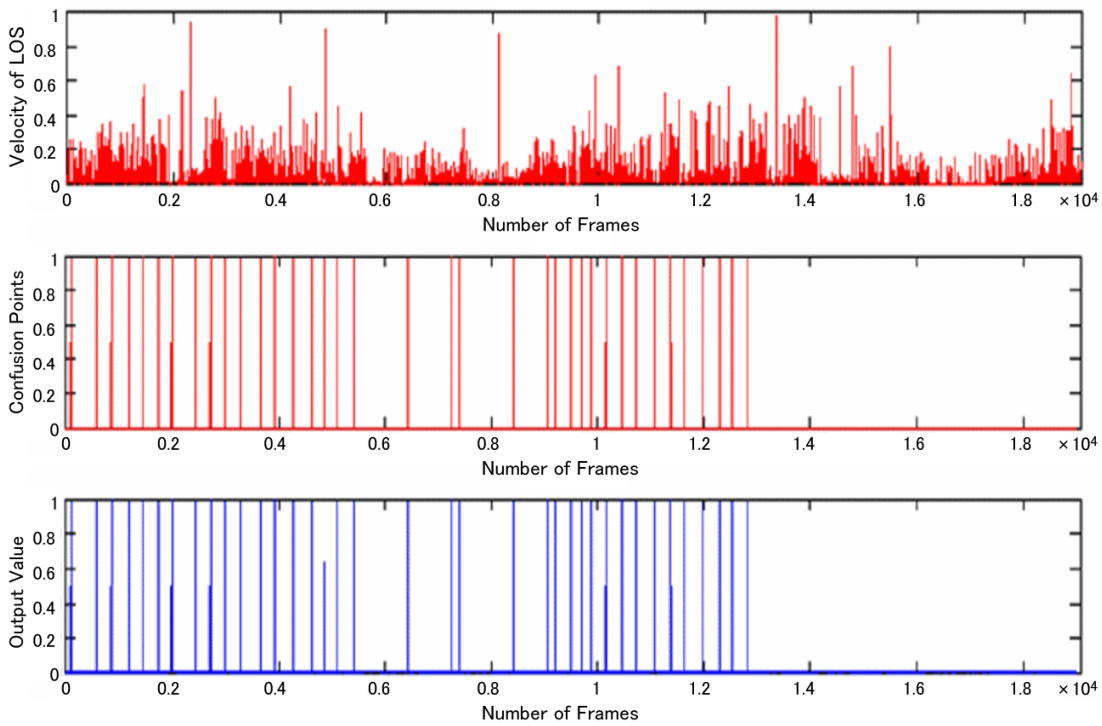

Figure 5. Verification Results for the Confusion Detection Model. The top chart shows the velocity changes with the lines of sight (LOS); the middle chart shows those confusion points where the model expected detection to occur; and the bottom chart shows the output values of the output units of the neural network.

Moreover, to verify the effectiveness of the model, we conducted another detection test using the noise-added data for further verification, including those data used for the neural network learning. The noise was calculated within a range of ±5% for each frame as a random value, and was then added to the original testing data. The detection rate of the confusion detection model with regard to such noise-added data was 100%.

2.2. Experiment 2

This experiment explored the possibility that a human state of confusion could be detected by a learning neural network using eye movement data limited to four kinds of stimuli in two key situations where subjects are contending with basic calculation problems, including a trap. The results of this experiment indicated that it was important to teach each neural network using the eye movement data recorded in each situation. Therefore, we tried to detect a state of confusion from eye movement data in a more realistic situation.

2.2.1. Experimental System

All factors in this experimental system were the same as in Experiment 1, except for the positioning of a video camera recorder with a microphone to film the display screen and simultaneously record the progress status of the task and the subject's voice. We used a Sony VX2000 (Sony Corp., Tokyo, Japan) video camera recorder.

2.2.2. Task and Procedure

The experiment focused upon using the Web and defined two situations to clarify states of confusion for Experiment 2, as described below:

(1) A state of confusion caused by not being able to find an item needed to accomplish one's goal; and

(2) A state of confusion caused by viewing a screen displaying an unexpected result and causing doubt as to whether or not one has committed an operational error.

As stimuli, two types of Web sites, each of which was constructed by drawing upon conventional Web services, were used. These stimuli were expected to provoke the two different states of confusion as described above. The type and details of each stimulus are shown in Table 4 and a sample screen shot for choosing a branch office in Task 5 is shown in Figure 6. The experimental procedure was modified for experiment 2 since it was more complex than the tasks in the experiment 1. In the experiment 1, the interviews were not conducted for all subjects. This decision was made from the point that no confirmation would be needed because the tasks were simple enough to understand. However, in the experiment 2, the interviews were considered to be a necessity for all subjects to confirm and to hold the reliability of the detected confusion because of the natural complexity of the task.

Task	Type	Detail
5	Online bank transfer	The subjects are asked to transfer a specified amount to a designated account displayed on the Web browser. However, when each subject selects the branch name, the system refuses to immediately display the target branch name on the screen (a type of time-lagged system). At this point, the subjects are confused.
6	Online hotel room reservation	The subjects are asked to make an online reservation on a Web browser in order to stay at a particular hotel with a friend. Each subject selects “two persons” on the third screen for the number of guests. However, the subjects will encounter incorrect output in the contents confirmation phase on the sixth screen, where the number of guests is displayed as seven persons. At this moment, each subject will start to harbour doubts about his/her own mistake, and will be confused.

Table 4. Type and Details of Each Experiment 2 Stimulus

Figure 6. Sample Screen of the Task 5 Shot. In this trap screen, the branch name choices are arranged in two columns and nine rows. However, the selected branch name cannot be found on the list.

The subjects of this experiment consisted of 44 males and 38 females. These subjects were both undergraduate and graduate students between the ages of 18 and 24 years old. All reported that they had normal vision, including corrected eyesight.

As was the case with Experiment 1, each subject carried out one randomly selected task. The number of subjects who carried out Task 5 was 31 and the number of subjects who carried out Task 6 was 51.

2.2.3. Detection Method

The detection algorithm was diverted from the algorithm that was used in Experiment 1. The neural network learning algorithm was the same used in Experiment 1. However, the neural network learning was conducted separately in the two tasks.

2.2.4. Detection Results

As a result of the neural network learning that used approximately 70% of all the data, the learning coalesced and a confusion detection model was created. To verify the detection accuracy of the model created, verification was conducted using the remaining approximate 30% of all the data as testing data. In addition, as a result of the recording during the experiment itself as well as the post-experiment interviews, the data revealed that several subjects had completed Task 6 without any awareness of the trap designed to cause confusion. These subjects' data were excluded from the neural network learning and from the detection rate verification. The post-verification detection rate was 100% in Task 5 and approximately 77.4% in Task 6.

2.3. Experiment 3

The results of Experiment 2 clearly indicate that a human state of confusion can be detected from eye movement data, even in realistic situations. Moreover, this shows that a human state of confusion can be detected by creating a model for each situation using eye movement data that were acquired from the respective situations. Consequently, an attempt was made to confirm whether age differences were a clearly indicated factor in the existence or lack thereof of a state of confusion. In order to verify the age difference variable, another test was conducted that contrasted the younger and older sets of test subjects from Experiment 3.

2.3.1. Experimental System

All components of this experimental system were the same as those used in Experiment 2, including use of a video camera recorder with a microphone.

2.3.2. Task and Procedure

In this experiment, Task 6, which was originally prepared as a stimulus for Experiment 2, was used because feedback and comments received in the post-Experiment 2 test subject interviews pointed out that Task 6 was easier to understand than Task 5.

The subjects of this experiment consisted of 62 older people, each with a minimum age of 65 years. However, among the data sets acquired, only 22 data sets were usable for objective analysis. This resulted from the eye movement data of older people being rather difficult to measure because of the diminished palpebra superior sizes caused by decreasing levator palpebrae superioris muscle tone. However, this particular subject group's computer and Web literacy skills did not pose any problem. The older test subjects belonged to a local computer club and possessed sufficient PC and Web navigation skills and knowledge.

2.3.3. Detection Method

The detection algorithm was diverted from the algorithm used in Experiment 1. The neural network learning algorithm was the same one used in Experiment 1.

2.3.4. Detection Results

As a result of the neural network learning that used approximately 70% of the total data, the learning coalesced and a confusion detection model was created. To confirm the detection accuracy of the model thus created, a verification of confusion was conducted using the remaining approximately 30% of all the data as testing data. In addition, it was confirmed that all subjects were confused at the trap point. The post-verification detection rate in Task 6 was 100%.

2.3.5. Confirming Age Differences

From the detection result of the model created using the data obtained from older people, it became clear that our proposed method had efficacy as a detection method for identifying the states of human confusion. Consequently, we also were able to confirm that the model built using the data obtained from young people was also reliable in detecting the state of confusion from the data of older people. After the confirmation, we checked as to whether the model created from the data of older people could also be applied to detect the state of confusion using the data obtained

from young people. Both the Task 6 model and data in Experiments 2 and 3 were used for this confirmation work.

The results showed that the detection rate of the young people model using the data of older people was 80%. In contrast, the detection rate of the older people model using the young people's data was only 50%.

3. Interest Estimation Method

Utilizing both the results of the development and the evaluation of our proposed method for detecting states of confusion, we confirmed that the internal state of humans could be identified by a model created to use learned neural network by velocity changes within lines of sight as the input data in a similar manner for the measurement of interest. That is, user interest in HCI could be estimated by converting the technique employed in the confusion detection method.

Experiment 4, also designed on the basis of our hypothesis, was an experiment conducted to collect eye movement data from subjects that showed interest in visual objects on screen monitors.

3.1. Experimental System

The experimental system used was the same as that employed in Experiment 1, except the use of a computer mouse was omitted.

3.2. Task and Procedure

Three types of tasks were prepared for this experiment. The tasks comprised a slide show that incorporated eight images, each of which was created using Microsoft PowerPoint 2003. During the process of each task, an image was presented for five seconds. The subjects were asked to view these tasks. The type and details of each task are shown in Table 5. The female subjects carried out all the tasks. However, the male subjects carried out only two of the three tasks (Tasks 7 and 8) because the subject of Task 9 was women's fashions.

Task	Type	Details
7	Mixed topic	Eight images from varied topics, such as beautiful scenery, actors/actresses, young animals, etc.
8	Automotive topic	Eight images pertaining to automotives. These images ranged from a station wagon, SUV, sedan, and so forth.
9	Fashion topic	Eight images pertaining to women's fashions. These images included different types of clothing, such as knitwear, long sleeves, a shirt, a coat, and others. These stimuli were used only for female test subjects.

Table 5: Types and Details of Each Experiment 4 Stimulus

The subjects were asked to evaluate their interest in the eight images after each task. Five containers and printed out images of the stimuli presented in the tasks were prepared. These were aligned alongside the experimental system in order to measure the eye movement data. The five containers were numbered one through five. These numbers indicated the intensity of interest. The specific evaluation procedure was as follows:

(1) The subjects were asked to select the one image that piqued their keenest interest from among the eight images.
(2) The subjects were asked to evaluate on a scale of one to five their intensity of interest in that single image and to put the image in the container.
(3) The subjects were asked to select the image that they thought had most keenly interested them from among the remaining seven images.
(4) The same evaluation sequence was repeated with the remaining six images and so on, until the last image had been evaluated.

The goal of this procedure was to identify the exact image that piqued the subject's keenest interest. Moreover, within the evaluation phase, this procedure made it possible to avoid the grading fluctuations that typically occurred when the images were sorted into numbered containers.

The subjects of this experiment consisted of 53 males and 46 females. They were all undergraduate and graduate students between the ages of 18 and 24. All reported that they had normal vision, including corrected eyesight.

3.3. Estimation Method

The data quantifying the velocity changes within the lines of sight was utilized as a parameter because the method used for detecting a state of confusion had been successful. The velocity changes within lines of sight were calculated in the same manner as the confusion detection method velocity changes within lines of sight had been calculated.

Firstly, to create an interest estimation model, the data were screened in accordance with the evaluation conditions. That is, did the subjects evaluate the images immediately or not? The data were then organized in order to check whether the first image chosen had piqued the subjects' keenest interests and thus were evaluated immediately. Those data where the subjects were more careful in making an evaluative decision were screened to check the possibility of whether the data included noise that could be relevant to tricking the subjects' imaginations. In that case, some data were not adopted for both the neural network learning and verification phases. The ultimate number relationship between the learning and the verification phases of the neural network is shown in Table 6.

	Number of Data	
	Learning	Verification
Task 7	25	57
Task 8	24	56
Task 9	9	31

Table 6. Breakdown of the Number of Data Used for the Neural Network Learning and Verification Phases

Secondly, in creating the interest estimation model, we endeavoured to use a learning neural network that employed those data quantifying velocity changes within lines of sight as the input data for extracting the features of a state of confusion when subjects exhibited interest in visual objects. The point where the subjects showed interest in visual objects was specified from the evaluation results assessed after collecting each task's eye movement data. The model created was used as a template,

and it includes features of interest for making estimations from eye movement data within time-series data.

3.4 Neural Network Learning Algorithm

The learning algorithm and the neural network construction were both the same as those used for the confusion detection method. The mass of input data was also the same but the learning policy differed. The labels used in the confusion detection method were modified. In this experiment, the input data mass was classified into two categories, the estimation point and a neutral point. The estimation point was set to start at 50 frames after the image was presented to the subject and it ended at 50 frames thereafter. The displayed image that was used to determine the estimation point's start was that image that the subject evaluated at the highest level. It was labelled as "1". This change was in keeping with the concept of time, which we measured in terms of the length of time that the subjects needed to take action after having acquired the visual information. In the confusion detection method, the preset time was based on the mental operation of Card et al. (1980). However, a person's interest was thought to require a longer process time than that of a user's confusion because the processing of the former type of information is more complex. This complexity is attributable to individual taste, a concept that itself formed over a period of many years in the field of psychological research. However, some uncertainty still remains. Details of the learning algorithm are shown in Figure 7. The input data were inputted one frame aside from the time-series data. During the learning process, the learning rate was set at 0.001 and the neural network ceased learning when the cost function value reached 0.00001. The neural network learned each piece of data that was acquired in each task.

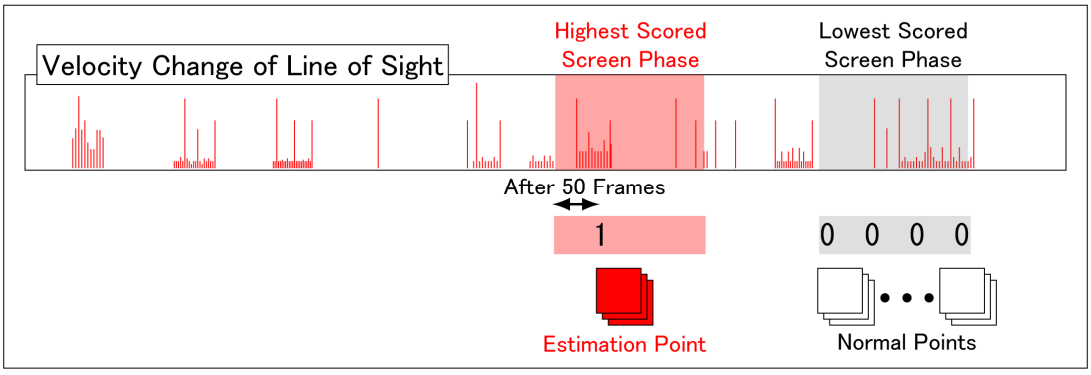

Figure 7. Learning Policy of the Experiment 4 Neural Networks

3.5 Estimation Results

As a result, the learning coalesced and a model to estimate a subject's interest was created. We confirmed the estimation ability by checking as to whether the model was able to estimate a subject's interest in a set of test data. In this study, we defined a condition of interest estimation as existing whenever the network output unit exceeded 0.5.

The estimation results of each model are shown in Table 7. In this verification, the data were not adopted as estimation objects, except for those data that were evaluated as the highest and lowest within a set of visual objects. For instance, when a subject accorded the highest valuation, 5 points, to visual object A, and accorded the lowest valuation, 1 point, to visual object B, and accorded 3 points to all the other visual objects, then, those visual objects accorded 3 points were not adopted as estimation objects. This chart reflects that, although the estimation point was only one per subject, all the data were merged into one.

	Estimation Rates [%]	
	Training Data	Testing Data
Task 7	72.0	64.9
Task 8	83.3	53.6
Task 9	100	83.9

Table 7. Estimation Results of the Neural Network

4. Discussion

This study showed that a method to detect a state of confusion, a human internal state, can be developed from eye movement data. This method demonstrated the importance of focusing upon a medium range history of eye movement data for approximately 50 frames.

In this paper, first we demonstrated the fact that a state of human confusion in simple HCI could be detected from eye movement data by focusing upon velocity changes within lines of sight as a parameter. This method used a learning neural network to employ eye movement data for pattern classification; those patterns became data labelled as confusion points and neutral points. The confusion points were those data

acquired at the time that subjects encountered calculated traps embedded in the experiment's tasks and the neutral points were those data acquired before subjects encountered such traps. The learning of the neural network coalesced and a confusion detection model was created. The model created was successful in verifying the remaining test data that had not been used for the neural network learning. The verification results demonstrated that the detection rates were satisfactory.

Second, we demonstrated that the states of human confusion during realistic HCI could be detected in a similar manner. Thus, the learning policy was modified from the detection method used for simple HCI. The learning coalesced, and we were able to obtain a high detection rate. These detection rates were comparable to other previous studies that endeavoured to recognize basic emotions utilizing data obtained from contact-free devices measuring biological information (Dellaert et al., 1996; De Silva & Ng, 2000; Nicolson et al., 2000; Umemuro & Yamashita, 2003).

Finally, an additional experiment was conducted, one that employed older people as subjects to verify whether age group differences appeared within the features of the states of confusion. As a result of the neural network learning replicating the procedure used to create the first model, the learning coalesced. The second model created had a high detection rate for testing data along with the models that were made from the young subject data. Moreover, in verifications conducted by age group, the young subject data model successfully detected the older subject data, and vice versa. Consequentially, the detection rates obtained from verifying the young subject data model against the older subject data was on average comparatively high. In contrast, the detection rate obtained from verifying the older subject data model against the young subject data was not demonstrably high. We found it apparent that there were effective eye movement differences between these two generations. For instance, the saccadic eye movement is known to be affected by the aging process (Abel, Troost, & Dell'Osso, 1983; Warabi, Kase, & Kato, 1984). However, because the aging effect upon eye movement characteristics varied greatly from individual to individual, the effect could not simply be quantified as maintaining a strict barrier between generations (Warabi et al., 1984; Moschner & Baloh, 1994).

From these results, it became clear that our proposed method provided high detection rates with a learning neural network that used velocity changes in lines of sight as input data, alongside data acquired from various test situations. This illustrates the possibility that a single model, one that integrates a sub-network created for more than one situation, can detect various states of confusion. Although the issue of the

age differences within this study still remains, we suspect that the problem lies not in the detection and algorithm but in the quality and quantity of our older test subject data. In addition, the knowledge gleaned in the verification by age group indicated the possibility that creating a single model could lead to data being detected without making any age or gender distinctions.

In order to establish a method for estimating human interest, we were able to create models displaying varying topical visual objects by employing almost the same algorithm as the one used in the confusion detection method. The human interest estimation models were also created with a learning neural network that used velocity changes in lines of sight as a parameter. However, the labelling process was modified based on the time-series differences pertinent to the internal states. In terms of a time-line, the state of confusion was thought to be the one next responsive in speed, following the state of situational reflection. In contrast, the time-line showing interest as an internal state was set at a longer time because humans require longer periods of time to make decisions. Therefore, we were able to demonstrate that human interest as an internal state can be estimated from eye movement data by focusing on the velocity changes in lines of sight. This method used a learning neural network to classify the labelled data of the estimation points and the neutral points according to patterns. The estimation points were those data measured at the time that the subject showed keen interest in the visual object and the neutral points were those data measured at the time that the subject exhibited absolutely no interest in the visual object. The neural network learning coalesced and the interest estimation models were thus created. As a result of those models' verification, these estimation models showed lower rates than the confusion detection rates were found to be in this study. The reason behind this is that there are differences in the fluctuating characteristics of each internal state, such as confusion and interest. For instance, in our research, confusion was a state that suddenly emerged from a normal situation. The difference between a normal and a confusing situation was clear and the difference likened to a digital waveform when the range of rise and fall of the state was delimited from “0” to “1”. In contrast, a state of interest always emerged more or less in the form of an analog waveform. Although we were able to intuitively understand these features, no precise index capable of measuring the characteristic of suddenness in an internal state exists. This feature highlights the difference of ambiguity. Therefore, we took into consideration the fact that the interest estimation results were lower than the confusion detection results.

Using these results, we found that estimating human interest from eye movement data at a higher accuracy rate required changing the learning algorithm and neural network policy. However, the higher estimation accuracy rate might prove as difficult to reach as a higher accuracy rate in the confusion detection method. This is because the fluctuating characteristics vary greatly between human interest and confusion.

Finally, an assumption was made that it was important to pay attention to the fluctuating characteristics of each internal state in order for our proposed method to be applicable to other internal states. Moreover, if an aim was made to identify more stable characteristics of internal states other than confusion, e.g., interest or intention, the identification rate should be expected to be lower, without hoping for a near 100% rate. In the case of identifying other fluctuating characteristics, appropriate changes need to be made to the target identification rate.

5. Conclusion

The results of these experiments supported our hypothesis that the changes in a person's internal state are somehow reflected in his/her eye movement data. The investigation results demonstrated that human internal states could indeed be identified by using eye movement data. These results lead us to the conclusion; it is highly possible that the states of human confusion and human interest can be identified by employing a neural network learning algorithm. In such an algorithm, the velocity changes within lines of sight are adopted to handle time-series data more quantitatively as eye movement history. Taking this knowledge into consideration, in order to identify human internal states, a collection of eye movement data is needed in each particular situation in conjunction with the learning of a neural network based on such collected data. This work shows some indication of the possibilities that a model can identify an internal state without reference to age or gender in a particular situation.

6. Acknowledgements

This study was supported by the Softopia Japan Foundation in Japan. We wish to thank Yasunori Tanaka, Masashi Nakamura, and the Densan System Co., Ltd., as well

as Dr. Minoru Sasaki and Dr. Yoshihiro Kaneko, both of Gifu University, for their many helpful suggestions during the course of this study.

7. References

Abel, L. A., Troost, B. T., & Dell'Osso, L. F. (1983). The effects of age on normal saccadic characteristics and their variability. *Vision Research, 23*, 33-37.

Baron-Cohen, S., Wheelwright, S., & Jolliffe, T. (1997). Is there a "language of the eyes?" Evidence from normal adults with autism or Asperger Syndrome. *Visual Cognition, 4,* 311-331.

Byrne, M. D., Anderson, J. R., Douglass, S., & Matessa, M. (1999). Eye Tracking the Visual Search of Click-down Menus. *Proceedings of the SIGCHI Conference of Human Factors in Computing Systems*, 402-409.

Card, S. K., Moran, T. P., & Newell, A. (1980). The keystroke-level model for user performance time with interactive systems. *Communications of the ACM, 23*, 396-410.

Card, S. K., Moran, T. P., & Newell, A. (1983). *The psychology of human-computer interaction.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Dellaert, F., Polzin, T., & Waibel, A. (1996). Recognizing Emotion in Speech. *Proceedings of the 4th International Conference on Spoken Language Processing*, 1970-1973.

De Silva, L. C., & Ng, P. C. (2000). Bimodal Emotion Recognition. *Proceedings of the 4th IEEE International Conference on Automatic Face and Gesture Recognition 2000*, 332-335.

Driver, J., Davis, G., Ricciardelli, P., Kidd, P., Maxwell, E., & Baron-Cohen, S. (1999). Gaze perception triggers reflexive visuospatial orienting. *Visual Cognition, 6*, 509-540.

Ekman, P. F. (1992). Facial expression of emotion: New findings, new questions. *Psychological Science, 3*, 34-38.

Hara, F., & Kobayashi, H. (1996). A Face Robot Able to Recognize and Produce Facial Expression. *Proceedings of the International Conference on Intelligent Robots and Systems,* 1600-1607.

Hayhoe, M., & Ballard, D. (2005). Eye movements in natural behavior. *Trends in Cognitive Science, 9*(4), 188-193.

Hess, E. H., & Polt, J. M. (1960). Pupil size as related to interest value of visual stimuli. *Science, 132,* 349-350.

Hess, E. H. (1965). Attitude and pupil size. *Scientific American, 212,* 46-54.

Ishino, K., & Hagiwara, M. (2003). A Feeling Estimation System Using a Simple Electroencephalograph, *Proceedings of the 2003 IEEE International Conference on Systems, Man and Cybernetics.* 4204-4209.

Izard, C. E. (1991). *The Psychology of Emotions.* N. Y.: Plenum Press.

Kobayashi, H., & Kohshima, S. (1997). Unique morphology of the human eye. *Nature*, *387*, 767-768.

Kobayashi, H., & Kohshima, S. (2001). Unique morphology of the human eye and its adaptive meaning: Comparative studies on external morphology of the primate eye. *Journal of Human Evolution, 40*, 419-435.

Lim, H., Ishii, A., & Takanishi, A. (2000). Emotion Expression of a Biped Personal Robot. *Proceedings of the 2000 IEEE/RSJ International Conference on Intelligent Robots and Systems, 1*, 191-196.

Lyons, M., Akamatsu, S, Kamachi M., & Gyoba, J. (1998). Coding facial expressions with Gabor wavelets. *IEEE International Conference on Automatic Face and Gesture Recognition*, 200-205.

Mace, R. L. (1985). Universal design: Barrier free environments for everyone. *Designers West, 33* (1), 147-152.

Macrae, C. N., Hood, B. M., Miline, A. B., Rowe, A. C., & Mason, M. F. (2002). Are you looking at me?: Eye gaze and person perception. *Psychological Science, 13,* 460-464.

Maglio, P. P., Barrett, R., Campbell, C. S., & Selker, C. (2000). SUITOR: An Attentive Information System. *Proceedings of the International Conference on Intelligent User Interfaces 2000*, 169-176.

Morris D. (1985). *Body Watching*. London: Equinox Ltd.

Moschner, C., & Baloh, R. W. (1994). Age-related changes in visual tracking. *Journal of Gerontology, 49*, M235-M238.

Musha, T., Terasaki, Y., Haque, H. A., & Ivanitsky, G. A. (1997). Feature extraction from EEGs associated with emotions. *Artificial Life Robotics, 1,* 15-19.

Nagamachi, M., Ito, K., & Tsuji, T. (1988). Image Technology Based on Knowledge Engineering and its Application to Design Consultation*. Proceedings of the 10th Ergonomics International, 88*, 72-74.

Nicolson, J., Takahashi, K., & Nakatsu, R. (2000). Emotion recognizing in speech using neural networks. *Neural Computing and Applications, 9* (4), 290-296.

Nielsen J. (1993). *Usability engineering*. San Diego, CA: Academic Press.

Norman, D. A., & Draper, S. W. (1986). *User centered system design: New perspective on human-computer interaction*. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Norman, D. A. (1988). *The psychology of everyday things*. New York: Basic Books.

Ohno, T. (2006). One-point calibration gaze tracking method, *Proceedings of the 2006 symposium on Eye Tracking Research & Applications*, 34.

Reeves, B., & Nass, C. (1996). *The media equation: How people treat computers, television, and new media like real people and places*. Cambridge, U.K.: Cambridge University Press.

Rumelhart, D. E., McClellans, J. L., & PDP Research Group. (1988). *Parallel distributed processing*. Cambridge: The MIT Press.

Takahashi, K. (2004). Comparison of emotion recognition method from bio-potential signals. *The Japanese Journal of Ergonomics*, *40* (2), 90-98.

Tomkins, S. S. (1962). *Affect, imagery, consciousness: Vol. 1. The positive affect*, N.Y.: Springer.

Tomkins, S. S. (1963). *Affect, imagery, consciousness: Vol. 2. The negative affect*, N.Y.: Springer

Umemuro, H., & Yamashita, J. (2003). Detection of user's confusion and surprise based on pupil dilation. *The Japanese Journal of Ergonomics, 39* (4), 153-161.

Warabi, T., Kase, M., & Kato, T. (1984). Effect of aging on accuracy of visually guided saccadic eye movement. *Annals of Neurology*, 40, 462-469.

Yacoob, Y., & Davis, L. S. (1996). Recognizing human facial expressions from long image sequences using optical flow. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *18* (6), 636-642.

Yamanobe, S., Taira, S., Morizono, T., & Kamio, T. (1990). Eye movement analysis system using computerized image recognition. *Archives of Otolaryngology and Head Neck Surgery, 116*, 338-341.

Zecca, M., Roccella, S., Carrozza, M. C., Miwa, H., Itoh, K., Cappiello, G., & Cabibihan, J., et al. (2004). On the Development of the Emotion Expression Humanoid Robot WE-4RII with RCHI-1. *Proceedings of the 4th IEEE/RAS International Conference on Humanoid Robots:, 1,* 235-252.

Zhu, Z., Fujimura, K., & Ji, Q. (2002). Real-time Eye Detection and Tracking Under Various Light Conditions. *Proceedings of the 2002 Symposium on Eye Tracking Research and Applications*, 139-144.