R&D of the Japanese Input Method using Life Log on an Eye-controlled Communication Device for Users with Disabilities

Kazuaki Shoji[1], Hiromi Watanabe[2] and Shinji Kotani[2]
[1]*Department of Education Interdisciplinary Graduate School of Medicine and Engineering, University of Yamanashi, Takeda 4-3-11 Kofu, Yamanashi, Japan*
[2]*Department of Research Interdisciplinary Graduate School of Medicine and Engineering, University of Yamanashi, Takeda 4-3-11 Kofu, Yamanashi, Japan*

Keywords: Japanese Input Method, Disability, Eye-control, Input Efficiency.

Abstract: We aim to enable the smooth communication of persons physically unable to speak. In our past study, we proposed three Japanese input methods using a portable eye- controlled communication device for users with conditions such as cerebral palsy or amyotrophic lateral sclerosis (ALS). However, these methods require nearly 30 seconds to cycle through one Japanese character. In this paper, we suggest a method to estimate the input word using the clues of nearby characters and accumulated experience. In addition, to raise the precision of the prediction, we use the connection between words based on a thesaurus. We have realized precise word conversion via a few input letters, as proved by the result of the simulation experiment.

1 INTRODUCTION

The number of handicapped people in Japan in 2006 was 3,570,000 (Ministry of Health, 2008). Among these, the number of physically handicapped people is 1,810,000 people. It is difficult for those with heavy cerebral palsy or later stages of ALS, to communicate by speaking, writing or gesturing.

Japan aims to create a society in which the physically disabled can take part in the economy and culture. It is not enough to simply improve the quality of life (QOL) for the physically disabled. Communication is also very important, but currently there is no effective method.

The final aim of this research is to achieve a symbiotic society of the disabled and non-disabled. Real-time communication system for disabled is necessary to achieve the society. Thus we aim to achieve via R&D the communication system.

This system needs to be customizable based on the nature of the disability and its severity. Moreover, cost and size of the system need to be considered, in order to make the system accessible to the greatest number of people. In addition, it is necessary to promote the system once it is created.

We use an eye-controlled communication device in a real-life system. The eye-controlled communication method is an interface, which can be used both with heavy cerebral palsy and with advanced ALS. Our final goals are developing and evaluating an eye-controlled communication device for a Japanese input method. We aim to use quantitative evaluation, qualitative evaluation by questionnaire and physiological evaluation with near-infrared spectroscopy (NIRS).

In the first stage, we established three Japanese input methods and evaluate these methods with NIRS (Kotani et al., 2010). In the second stage, we will improve the system based on feedback from users. In the third stage, we will develop the new system using cellular phones and their digital cameras. There has already been a great deal of research of eye-controlled communication devices (Hori and Saitoh, 2006; Yamamoto, Nagamatsu and Watanabe, 2009), with some products on the market. There has even been research on the Japanese touch screen input method (Noda et al., 2008). However, an effective Japanese input system using eye-controlled communication has not been developed.

In our previous studies, we have succeeded in developing three Japanese input methods that are not burdensome to users. These methods encountered we had a problem that the users with heavy cerebral require to take approximately 30 seconds to enter one

Shoji K., Watanabe H. and Kotani S..
R&D of the Japanese Input Method using Life Log on an Eye-controlled Communication Device for Users with Disabilities.
DOI: 10.5220/0004681202070213
In *Proceedings of the International Conference on Physiological Computing Systems* (PhyCS-2014), pages 207-213
ISBN: 978-989-758-006-2

Japanese character.

This paper proposes a predictive conversion system (PCS) for high efficiency character input. The system uses various information accumulated for the user, neighboring information, and information regarding time and place. The new system will enable smooth communication of the physically disabled, which will contribute enormously to the rest of the society. It is also expected that the study will help improve existing input system for the able-bodied.

This study was approved by the institutional ethics committee of University of Yamanashi.

2 JAPANESE INPUT SYSTEM

2.1 Eye-controlled Communication Device

My Tobii P10 is a portable eye-controlled communication device which is made by Tobii Technology Inc. Everything, including a 15" screen, eye control device and computer, is integrated into one unit. We use this unit in the first stage of our research.

My Tobii P10 has sold more than 2,000 units. It received excellent reviews as a user-friendly, eye-controlled communication device. Unfortunately, the device does not have a Japanese capability.

We proposed three Japanese input methods: (1) "Roman letters" on-screen keyboard, (2) "Japanese kana" input, and (3) "modified beeper" input. The "modified beeper" omits numbers. Figure 1 shows the "Roman letters" on-screen keyboard image, Figure 2 shows the "Japanese kana" input image and Figure 3 shows "modified beeper" input image.

2.2 NIRS

Pocket NIRS Duo is a portable near infrared spectroscopic device which is made by Hamamatsu Photonics. The spectroscopic method (Villringer et al., 1993) is a non-invasive imaging of brain function. The technique is safe and has lower constraints for subjects compared to other brain imaging techniques. In addition, only a simple and easy-to-use system is needed for the measurements.

The positions of the probes are Fp1 and Fp2, EEG (Electroencephalogram) measurements according to the 10-20 international electrode placement system (Okamoto et al.,2004; Moosmann et al., 2003, Roche-Labarbe et al., 2008).

2.3 Evaluation

We evaluated three Japanese input methods, including quantitative, qualitative evaluation and physiological evaluation with NIRS system.

We observed and evaluated four able-bodied Japanese speakers using the system, six times every two weeks for three months. The results of proficiency evaluation showed the Roman-letter input method was most effective (Kotani et al., 2011). It had the highest rate of input characters for all the subjects. One subject was able to double the input rate over time. These results show the effectiveness of the system for subjects with light cerebral palsy or ALS.

We list the disabled subjects in Table 1 below.

Table 1: Detailed information of the 2 disabled subjects.

Subjects	Age	Sex	Conditions
HS01	teenage	Female	Heavy cerebral palsy
HS02	twenties	Female	Heavy cerebral palsy

Figure 1: "Roman letters" on-screen keyboard image.

Figure 2: "Japanese kana" input image.

Figure 3: "Modified beeper" input image.

The frequency of the experiment was every month for HS02 and every two weeks for HS01. HS01 could input a symbol after 4 months. She could input

her name after 5 months. HS02 could input her name and family name after two months. She could input her favorite lyrics in four months, and after that she could control Windows Media player with her glance in six months.

HS01 and HS02 used the modified beeper. The users with heavy cerebral palsy could not use the Roman-letter input method, because they could not fix their glance to a small area. For this reason, character input efficiency decreased.

The users with heavy cerebral palsy require approximately 30 seconds to enter one Japanese character. To solve this problem, we studied an efficient predictive conversion system.

3 PREDICTIVE CONVERSION METHOD

This system is proposed for the Japanese input system, but to facilitate explanation in English, we will use English words.

3.1 Summary

The main features of the predictive system are as follows:

i. Raise the priority of the word that had been input in the past.
ii. A user picks a dictionary.
iii. The system also uses the context.

Because past input decision data did not exist for the users in this study, an appropriate dictionary did not exist. To make a word database, we fused static attributes from the life log, the basic attributes by time and place, and dynamic attributes by the situation and target estimation. The word database improves prediction conversion precision. We show each item in Table 2.

Table 2: Detailed information of the attributes.

Attribute	Item	Contents
Static attributes	Knowledge	Memorize person
	Experiment	Memorize proper nouns
Basic attributes	Time information	For choose greeting
	Place information	For choose topic
Dynamic attributes	Situation	Sound recognition
	Target estimation	Object recognition

Below we give examples of situations which would all result in the letter “g.”

“Good morning”: from basic attributes, 8:00 A.M.

“Good night”: from basic attributes, 22:00 P.M.

“George”: from static and dynamic attributes, the person being addressed is George.

“Garden”: from the basic attribute of being in a garden.

“Glad”: from dynamic attributes, when somebody says “Congratulations.”

“Grape”: from the dynamic attribute of having a grape in front.

The system can thus estimate a right word.

In the communication of a disabled person, it is very difficult to perform a word estimate in an accident. However, word estimates are possible.

3.2 System Configuration

We show each configuration item in Figure 4.

Static attribute information is formed by visual information provided by a camera and sound collected by a digital voice recorder. The system creates a life log from proper nouns, showing more frequent use than the accumulated image and sound. The word database is currently created by manual operation. For higher efficiency dictionary generation, we can use the method that automatically analyzes a proper noun (Coates-Stephens, 1993).

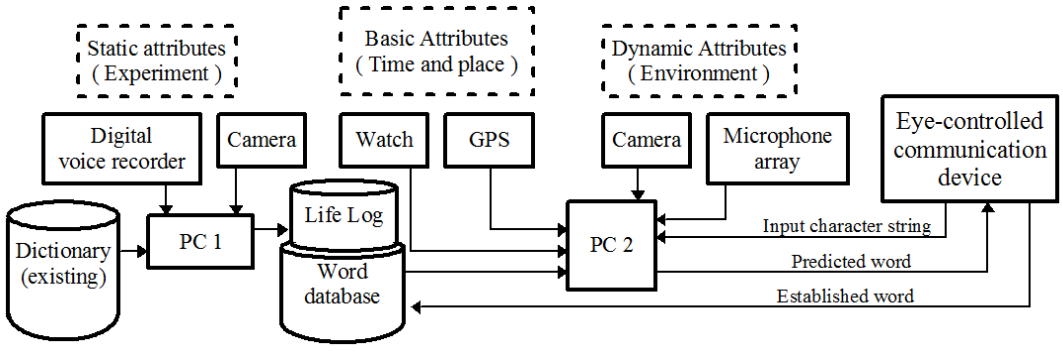


Figure 4: General view of the system.

To get basic attribute information, the system acquires the time from a clock with a date and time and acquires the position from GPS. A word estimate corresponding to time and place is useful for the input of the basic words, such as greetings.

The system acquires neighboring image information and sound information and uses it as a dynamic attribute. The system uses general object recognition (Zhang, Zelinsky and Samaras, 2007) to identify images. Because physically disabled persons often discuss objects in the vicinity, the system can guess words based on the object. The system identifies the person by using facial and body recognition (Lanitis, Taylor and Cootes, 1996). The system compares these results with the static attributes. The system acquires sound via a microphone array to grasp the topic that a person in the vicinity is talking about. The acquired sound then undergoes source localization via the HARK or MUSIC methods (Nakadai et al., 2010). This enables sound recognition that is robust against multiple sources of noise. Furthermore, words can be guessed from the recorded sound using Julius/Julian (Lee et al., 2001).

4 WORD DATABASE

4.1 Features

We use the connection between words based on a thesaurus (Chen et al., 2003) used by a search system. We defined the connection as a word tree.

The system has a plural word tree built-in. Each word tree has a condition that activates it. When a word tree has been activated, the input system gives priority to words in the word tree and predicts them. For example, there is a word tree, which has "apple pie" and "baked apple" under "apple," and "apple" under "fruit." The word tree is activated when one of the conditions shown below is satisfied.

The system is currently located in a greengrocer's.

A user inputs the word "food" or "eat" by eye-controlled communication device.

The system recognized the word "food" or "eat" from sound information.

These are the conditions that are met when a topic is likely to be "food" and "fruit." When either condition was met, or the user inputs "a," "apple" takes high priority and is chosen as the first candidate word.

We show the advantages of this word tree in Figure 5. the system can treat "apple" as food as separate from "apple" as a blossom. Therefore, the system can prevent "baked apple" being given priority when the topic is blossoms.

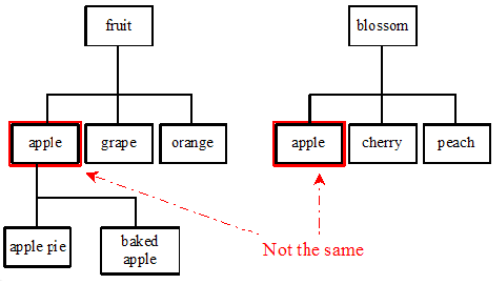


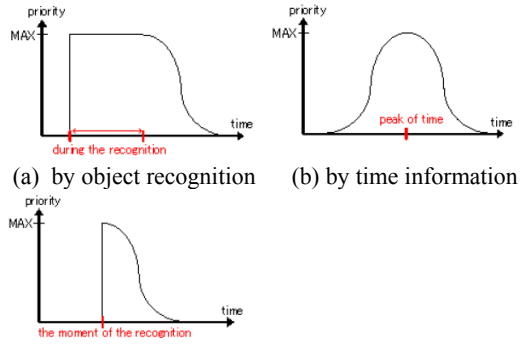
Figure 5: Feature of word tree.

4.2 Active and Inactivate

At any given time, each word tree has a priority, which changes with time. The priority changes with Expression 1, modeled on a Gaussian. In the expression below, *MAX* indicates maximum priority, t is elapsed time, and σ is the rate of decrease:

$$priority = MAX \exp\left(-\frac{t^2}{2\sigma^2}\right) \quad (1)$$

The predicted change depends on time information, as well as object and sound recognition. When the word tree is activated by time information, priority reaches a maximum. We show the rough shape of the priority in Figure 6.



(a) by object recognition (b) by time information

(c) by sound recognition

Figure 6: The shape of change in word tree.

When a tree is activated by object recognition and sound recognition, the moment of activation is the peak of the priority, subsequently decreasing with time. We show the rough shape of the priority in such case in Figures 6 (a)-(c).

The words are then displayed as predictive input candidates, starting with trees with the highest priority. When multiple trees are activated, if there

are synonyms among them, the priority of the words with synonyms increases the priority of their trees.

4.3 Implementation Approach

We write the word trees in the XML format that is currently the best in expressing hierarchical structure. We show the example describing the fruit tree of Figure 5 in Figure 7. Words are denoted by <word> in the hierarchical structure.

```
<word tree>
  <word name= "fruit">
    <trigger  type="snd"  ,obj  ="eat",
max="1.0", reduce="0.5">
    <word name="apple">
      <word name="apple pie"></word>
      <word name="baked apple"></word>
    </word>
    <word name="grape"></word>
    <word name="orange"></word>
  </word>
</word tree>
```

Figure 7: Example of an XML description of word tree.

The activated condition is described in the <trigger> element. The condition, the maximum of priority and the rate of decline of priority are described in attributes. To take an example of object recognition, "snd" is the attribute type. In addition, the recognized contents are described under "eat," and the maximum value of the priority is described in max, and a rate of decline is set in reduce.

The system reads this XML when it boots up. A word tree is activated when the activation conditions in the XML matches the information provided by each module.

4.4 Theoretical Evaluation

Below we give an example: consider the two word trees in Figure 8. There are three cases: (1) when neither word tree is active, (2) when only the "fruit" is active, (3) when both trees are active. In each case, we compared the priority of the word "grape." We registered 10, 20 or 30 other words that begin with "g" in the dictionary. For words whose trees' priority remains constant, predictions are given in alphabetical order.

The results are presented in Figure 9. We show the number of words that begin with "g" in the dictionary on the horizontal axis, and the order in which "grape" appears on the vertical axis. When neither word tree is activated, there is a correlation between the order of "grape" and the number of registered words. However, when the "fruit" word

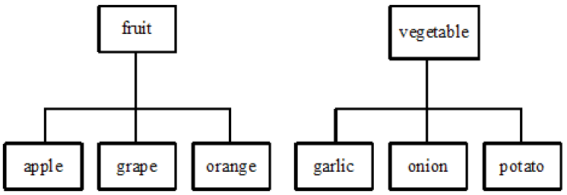


Figure 8: Two related word tree.

tree is activated, "grape" comes first. When both word trees are activated, because "garlic" is first alphabetically, "grape" comes second.

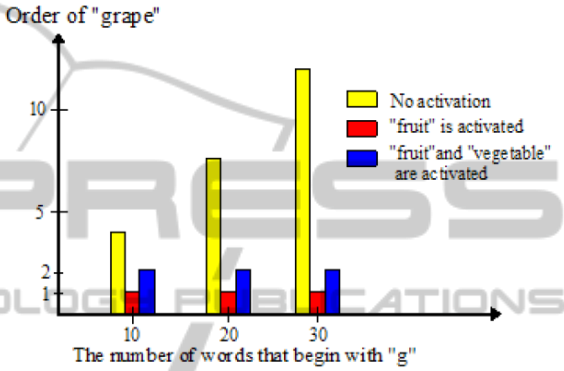


Figure 9: The result of theoretical evaluation.

5 EXPERIMENT

This experiment was carried out to test the efficiency of using word trees. We show the results of a simulation, because the accumulation of life log data is insufficient for these purposes.

5.1 Experimental Method

This experiment assumed the input of "good morning" by predictive conversion with using word trees. "Good morning" is included in the word tree whose priority becomes maximum at 8:00 A.M. During an experiment, time information is changed at random from 6:00 A.M. to 10:00 A.M. The priority of "good morning" has the shape of a Gaussian, changing from 0.6 to 1.0 at random.

A number of words to begin in "g" shall be activated. We assume that the priority of these words changes from 0.0 to 1.0 at random. We examined the relationship between the order of "good morning" and the number of activated words. We tried each pattern 500 times.

5.2 Experimental Result

The results of the simulation are presented in Figure 10. The horizontal axis indicates the number of words these are on the activated word tree. The vertical axis shows the order of "good morning" and its probability.

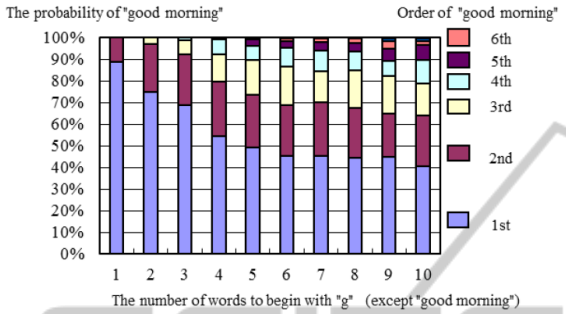


Figure 10: The result of experiment.

Figure 10 tells us that the probability that "good morning" comes first even if ten other words beginning with "g" are activated exceeds 40%.

In the situation of real-life use, we are likely to input "good afternoon," "good evening," and "good night" the day before we input "good morning." Therefore, the priority of "good morning" lowers by the conventional predictive method that uses recent input information in the past. When we use the prediction using the word tree, the probability that the order of "good morning" comes after the fourth place, even if there are ten other activated words beginning with "g," is around 20%. From this viewpoint we can say that input efficiency improves in comparison with the conventional predictive method.

When its priority is in the fourth place, "good morning" is more likely to be predicted if we input "go." When we use conventional predictive method, "good afternoon" "good evening" and "good night" are more likely to be given priority, even if we input "good." This example makes it clear that the amount of input necessary for the correct word decision can decrease with the new system.

In this experiment, we set the priority of the other words beginning with "g" at random. It will be necessary to consider an effective priority change using the life log and dynamic attributes in the future.

6 CONCLUSIONS

In this paper, we have suggested a novel method to improve input efficiency that has plagued eye-controlled input systems. We demonstrate a new predictive conversion method, which utilizes both the life log of the user and his/her environment. This method enables educated guesses in a variety of situations using static, basic and dynamic attributes. We use data from various sensors, such as a camera and a microphone array, to implement this new system.

We furthermore propose the use of word trees to make predictions more effective. The word tree is a way of thinking based on the thesaurus, and it can be described by XML. It is generally believed that the cost of the construction of all of the relevant word trees may be enormous. Through the use of automatic tree-building of a thesaurus (Kageura, Aizawa and Tsuji, 2000; Chen et al., 1995), this problem is solved.

Our simulation showed that we have indeed obtained efficient prediction using word trees. Multiple word trees were activated; even if in a complicated situation, we made more efficient predictions than the conventional method.

For now, only the dynamic attribute acquisition using sound has been implemented. We aim to suggest a method of general object recognition using images in the future. For this implementation, we are studying a method using depth information to be provided by a personal detection Kinect sensor for the Xbox360. We also create a life log by general object recognition and sound recognition. We plan to evaluate the resulting system using the additional sources of information for making predictions.

REFERENCES

Chen, H., Schatz, B, R., Yim, T. and Fye, D., 1995, Automatic thesaurus generation for an electronic community system, *Journal of the American Society for Information Science (JASIS)*, 46(3), pp.175-193.

Chen, Z., Liu, S., Wenyin, L., Pu, G. and Ma, W,Y., 2003, Building a Web Thesaurus from Web Link Structure, *Proc. of the ACM SIGIR*, pp.48-55.

Coates-Stephens, S., 1993, The analysis and acquisition of proper names for the understanding of free text, *In Computers and the Humanities*, Vol.26, pp.441-456.

Hori, J. and Saitoh, Y., 2006, Development of a Communication Support Device Controlled by Eye Movements and Voluntary Eye Blink, *IEICE TRANS. INF.&SYST.*, vol.E89-D, no.6, pp.1790-1797.

Kageura,K., Tsuji, K. and Aizawa, A., 2000, Automatic thesaurus generation through multiple filtering, *Proc. of the 18th International Conference on Computational Linguistics*, pp.397-403.

Kotani, S. Ohgi, K., Watanabe, H., Komasaki, T. and Yamamoto, Y., 2010, R&D of the Japanese Input

Method using an eye-controlled communication device for users with disabilities and evaluation with NIRS, *Proc. 2010 IEEE International Confrence on Systems, Man, and cybernetics (SMC)*, Isutanbul, pp.2545-2550.

Kotani, S., Tanzawa, T., Watanabe, H., Ohgi, K., Komasaki, T. And Kenmotsu, T., 2011, Proficiency evaluation of three Japanese input methods using an eye-Controlled communication device for users with disabilities, *Proc. 2011 IEEE International Conference on Systems, Man, and cybernetics (SMC2011)*, Alaska, pp.3230-3235.

Lanitis, A., Taylor, C, J. and Cootes, T, F., 1996, An automatic face identification system using flexible appearance models, *Image and Vision Computing*, vol.13, no.5, pp.393-401.

Lee, A., Kawahara, T. and Shikano, K., 2001, Julius -- an open source real-time large vocabulary recognition engine, *Proc. EUROSPEECH*, pp.1691-1694.

Ministry of Health, Labour and Welfare., 2008. Statics data. in Japanese.

Moosmann, M,. Ritter, P,. Krastel, I,. Brink, A., Thees, S., Blankenburg, F., Taskin, B., Obrig, H. and Villringer, A., 2003, Correlates of alpha rhythm in functional magnetic resonance imaging and near infrared spectroscopy, *NeuroImage*, 20, pp.145-158.

Nakadai, K., Takahashi, T., Okuno, H, G., Nakajima, H., Hasegawa, Y. and Tsujino, H., 2010, Design and Implementation of Robot Audition System "HARK", *Advanced Robotics*, vol.24, no.5-6, pp.739-761.

Noda, T., Shirai, H., Kuroiwa, J., Odaka, T. and Ogura, H., 2008, The Japanese Input Method Base on the Example for a personal Digital Assistant, *Memories of the Graduate School of Engineering*, University of Fukui, 56, pp.69-76.

Okamoto, M., Dan, H., Sakamoto, K., Takeo, K., Shimizu, K., Kohno, S., Oda, I., Isobe, S., Suzuki, T., Kohyama, K. and Dan, I., 2004, Three-dimensional probabilistic anatomical cranio-cerebral correlation via the international 10-20 system oriented for transcranial functional brain mapping, *NeuroImage*, 21, pp.99-111.

Roche-Labarbe, N., Zaaimi, B., Berquin, P., Nehlig, A., Grebe, R. and Wallois F.,2008, NIRS-measured oxy- and deoxyhemoglobin changes associated with EEG spike-and wave discharges in children, *Spilepsia*, 49(11), pp.1871-1880.

Villringer, A., Planck, J., Hock, C., Schleinkofer, L. and Dirnagl, U., 1993, Nera infrared spectroscopy (NIRS): A new tool to study hemodynamic changes during activation of brain function in human adults, *Neuroscience Letters*, vol.154, Issues 1-2, pp.101-104.

Yamamoto, M., Nagamatsu, T. And Watanabe, T., 2009, Development of an Eye-Tracking Pen Display based on the Stereo Bright Pupil Technique, *Technical report of IEICE. HCS*, 109(27), pp.147-150.

Zhang, W., Zelinsky,G. and Samaras, D., 2007, Real-Time Accurate Object Detection Using Multiple Resolutions, *Proc. IEEE Intenational Conference Computer Vision*, pp. 1-8, 2007.