Usability analysis of home electrical appliances based on eye tracking and physiological data

Ryoko Fukuda, *Member, ISG*

***Abstract*—Gerontechnology products may help daily lives of older adults. However, "useful home electrical appliances" are sometimes so complicated that older adults have great difficulties with using them. In order to identify difficulties for older adults, use of simulated rice cookers was analyzed by behavior observation, eye tracking, physiological measurement, and interview. The results showed that complicated function allocation and ambiguous labeling of operation keys confused the older users. In order to avoid such difficulties, definite classification of functions and explicit labeling are needed.**

I. Introduction

Home electrical appliances have improved our life by reduction of work in housekeeping. It should also be true for older adults, however it is not always so. Some older adults have bad experiences with or fears about household technologies and need simplification of use [1]. Modern home electrical appliances provide us various "useful" functions. For instance, a microwave range is often combined with an electric oven and various pre-defined cooking menus are available with help of computer. A modern rice cooker cooks not only rice, can also make yoghurt and bake bread. However such multifunctional home electrical appliances are sometimes so complicated that only a part of the functions are actually utilized. As a result, above mentioned problems become more serious, especially for older users who are not familiar with such modern "high-tech" products.

Difficulties for older adults are mainly related with age-related declines. Human characteristics which should be considered there can be classified into three factors: sensation/perception, cognition, and movement control [2]. Of those, cognition is strongly related to use of high-tech products including above mentioned multi functional modern home electrical appliances. In order to find and use the aimed function, users should understand the structure of function allocation in the user interface and the operation flow. In addition, the attitude of older adults toward high-tech products affects also performance of their use.

For the improvement of high-tech products regarding older users, it should be clarified which kinds of cognitive problems are especially serious for older users. Conventionally subjective evaluation methods based on a questionnaire or an interview have been applied for that purpose. They provide surely important information about difficulties for older users, but sometimes there is a limit. If older adults have some difficulties, they blame their difficulties not on an inconvenient design of a product but on themselves, because they are aware of their age-related declines. As a result, a subjective evaluation by elderly users can be disproportionately good [3]. In order to avoid this problem, objective evaluation methods should also be employed. Already some experimental studies have pointed out that physiological data show significant sensitivity to changes in stress levels and can bring important information for usability testing [4][5]. Recently subjective evaluation, physiological measurement, and behavior observation have been employed in combination for usability testing [6]. Although there are some limitations, objective evaluation based on physiological measurement can contribute for the identification of cognitive difficulties related to usability. In addition, eye tracking data is frequently applied for the analysis of user interface [7].

Therefore, this study aimed to clarify difficulties for older users in using multifunctional home appliances by behavior observation, eye tracking, physiological measurement, and interview.

II. Methods

A. Subjects

Seven subjects (four males and three females, age: 59–71) took part in this experiment. The data of two subjects had not enough quality to analyze, so that the data of five subjects were employed for the analysis.

B. Stimuli and apparatus

User interfaces of five real rice cookers were simulated by using a touch panel display. For all user interfaces, same fonts and background colors were applied. Table 1 shows the detailed features of each stimulus. In order to cook rice properly, the appropriate cooking menu should be selected regarding type of rice and texture of cooked rice. Menus for mixed rice and porridge are also provided by all the user interfaces employed in this experiment. However they were allocated differently to menu keys. A part of the stimuli offered also menus other than rice recipes.

Eye movements of subjects were recorded by the eye tracking system EMR-8B (Nac Image Technology, Inc.). Data sampling rate was 30 Hz and spatial resolution was better than 0.1 degree visual angle. For the physiological

Manuscript received January 31, 2008. This work was supported in part by 'High-Tech Research Center' Project for Private Universities' matching fund subsidy from MEXT (Ministry of Education, Culture, Sports, Science and Technology), 2005-2009.

Ryoko Fukuda is with Faculty of Environment and Information Studies, Keio University, Fujisawa, Kanagawa, 252-8520 Japan (phone: +81-466-49-3525; fax: +81-466-47-5041; e-mail: ryoko@sfc.keio.ac.jp).

Table 1. Applied stimuli

Stimulus A

Menu keys: (1) "Rice" (type of rice) (2) "Hardness" (texture of cooked rice) (3) "Menu"
Key to start timer setting: (4) "Easy programming"
Time setting keys: (5) "▲" (reverse)
(6) "▼" (advance)
Key to activate timer: (7) "Rice cooking"

Stimulus B

Menu keys: (1) "Rice cooking menu" (2) "Easy menu" (3) "Hardness" (texture of cooked rice)
Key to start timer set: (4) "Programming"
Time setting keys: (5) "Hour" (6) "Minute"
Key to activate timer: (4) "Programming"

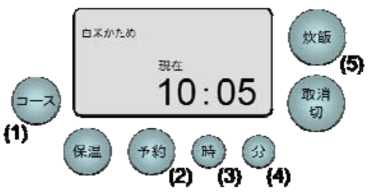
Stimulus C

Menu keys: (1) "Menu" (2) "Cooking type" (texture of cooked rice) (3) "Course"
Key to start timer set: (4) "Programming"
Time setting keys: (5) "Hour" (6) "Minute"
Key to activate timer: (7) "Rice cooking"

Stimulus D

Menu keys: (1) "Healthy menu" (2) "Rice cooking menu" (3) "Pre-washed rice"
Key to start timer set: (4) "Programming"
Time setting keys: (5) "▼" (advance)
(6) "▲" (reverse)
Key to activate timer: (7) "Rice cooking"

Stimulus E



Menu key: (1) "Course"
Key to start timer set: (2) "Programming"
Time setting keys: (3) "Hour" (4) "Minute"
Key to activate timer: (5) "Rice cooking"

measurement, PowerLab data acquisition system (ADInstruments) and Chart software were employed. In order to measure pulse, respiration, and galvanic skin response (GSR), this system was equipped with a piezo electric pulse transducer (MLT1010), a piezo respiratory belt transducer (MLT1132), a GSR amp (ML116), and a GSR finger electrodes (MLT116F). Data sampling rate was 1000 Hz. Eye tracking video and physiological data was synchronized by capturing eye tracking video into Chart software. Behavior of subjects was recorded by a video camera (Fig. 1).

Fig. 1. Experimental setup

C. *Procedure*

After calibration of the eye tracking system and the data acquisition system, the subjects were asked to select certain cooking menu and set the timer by using one of the simulated user interfaces of real rice cookers. During task accomplishment, behavior, eye movements, physiological data of subjects were recorded. After each trial, the subjects were interviewed about their difficulties in the task accomplishment. All subjects accomplished five trials and order of stimulus was differed among subjects.

III. Results

A. *Required time and number of operation for task accomplishment*

In order to select a cooking menu, 10 – 30 s were required on the average. The setting of timer required far longer time; in the shortest case it took only 13 s, whereas longer than 2 min in the worst case. On the average, Stimulus A required the longest time. Both in the time for selection of cooking menu and that for timer setting, a significant difference among stimuli was confirmed (ANOVA, selection of cooking: $p < .01$, timer setting: $p < .05$).

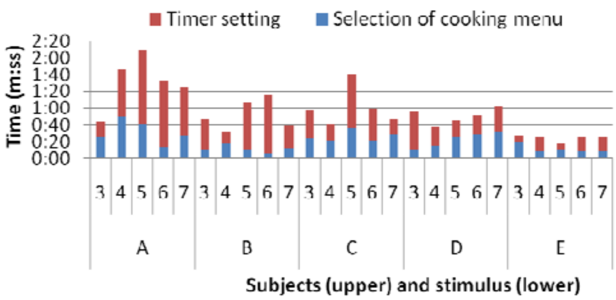
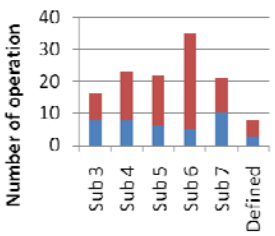
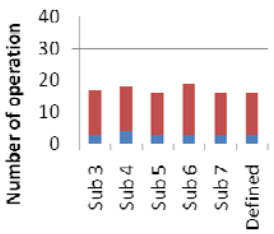


Fig. 2. Required time for task accomplishment

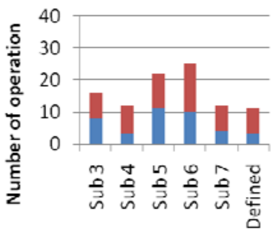
Above mentioned difference in required time for task accomplishment was related to the required number of operation. Therefore the number of the defined operation steps and observed number of operation were summarized in Fig. 3. With regard to Stimulus B and E, subjects required not many extra operations. Some subjects made no extra operation than the defined operation steps. On the other hand, Stimulus A, C, and D, the observed number of operation was far more than that in the defined operation steps. Especially for the timer setting of Stimulus A, the subjects have done 16 operations on the average where actually only five operations were required.



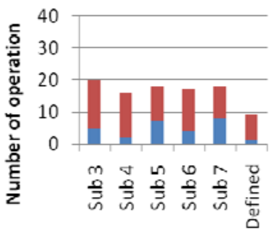
(a) Stimulus A



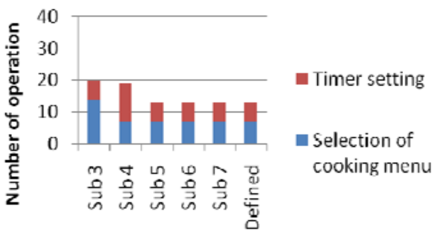
(b) Stimulus B



(c) Stimulus C



(d) Stimulus D



(e) Stimulus E

Fig. 3. Observed number of operation

B. *Observed errors and difficulties*

In most of the trials, subjects operated in a different way from the system designers' intention.

One of the most typical errors was the selection of a wrong key. If there were two or more similar menu keys, subjects wondered which key was the aimed one. For example, Stimulus B provided "Cooking menu" key and "Easy menu" key. In Stimulus C "Menu" key, "Cooking type" key, and "Course" key were available. With regard to these stimuli, the subjects took longer time to select one menu key at first. After selection of a wrong key, the subjects required extra operation steps and time.

Another error was observed during selection of the aimed cooking menu. After choosing one menu key, the subjects pressed that key to switch the menu one by one. In some cases, the aimed menu was passed. It is supposed because subjects paid not much attention to the display or they pressed the key rhythmically. In such cases selection process had to be repeated again. Similar error was also observed in time setting.

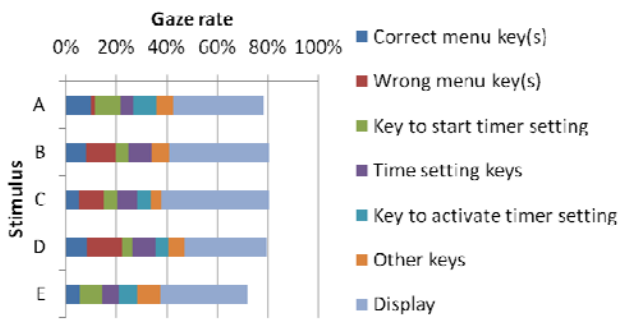
As for time setting, keys labeled with "▼" and "▲" in Stimulus A and D caused some difficulties. Each triangle indicated the direction of time transition produced by key press. But the meaning was not consistent between two stimuli: "▲" meant set back the clock and "▼" meant to advance the clock in Stimulus A, whereas the meanings of these signs were inverse in Stimulus D. The layout of two keys was also different: two keys were arranged vertically in Stimulus A, whereas they were arranged horizontally in Stimulus D. These differences confused the subjects. In Stimulus A, three subjects pressed the wrong key at first in order to set back the time. Other one subject pressed the correct key first, but then this subject pressed the wrong key. In Stimulus D, three subjects selected the correct key and two subjects pressed the wrong key. As for the stimuli provided two buttons labeled "Hour" and "Minute", only one subject had made extra button press.

After setting the time, all the stimuli required to press one more key in order to activate the timer. Especially in the first trial, subjects did not think of this process and thought that they have already completed the given task. After the additional instruction was given, subjects searched for the required key. Except Stimulus B, "Rice cooking" key was required. Only in Stimulus B, the timer was activated by "Programming" key. One subject pressed "Rice cooking menu" key instead of "Programming" key.

C. *Eye tracking data*

In the analysis, areas of interests (AOIs) were defined for each stimulus and the gazed AOI was identified frame by frame.

Overall, similar keys were repeatedly compared before one of them was selected. Even if the wrong menu key was selected, all the menu keys were compared again during selecting the menu key for the further time. Fig. 4 shows the gaze rate on individual AOIs in each stimulus. Gaze rate on menu keys were lower in Stimulus E which provided only one menu key. Of the four other stimuli included three menu keys each, total gaze rate on menu keys in Stimulus A was lower than other stimuli.



*In Stimulus B, the key to start timer setting and the key to activate timer setting was same.

Fig. 4. Gaze distribution

Only in Stimulus C, the timer was activated by pressing "Programming" key instead of "Rice cooking" key. During searching for the key to active timer, the right upper part was frequently gazed, where "Rice cooking" key was provided in other four stimuli.

During operation, the display was mainly looked. If a key press brought changes in two indicators, sometimes only one change was regarded for the first time. The gaze at other changed indicator was delayed in the early phase of task accomplishment. Eye tracking data during pressing the menu key or time setting keys revealed that passing the aimed menu or time was caused by rhythmical key press. Although the subjects gazed at the indicator which showed that the aimed menu or time was selected, they pressed the key further.

D. Relationship between physiological data and behavior

Pulse data and respiration data were converted into number of beat per minute (BPM). However, they varied not much during trial. On the contrary, the variation of GSR data was clear to observe and also immediate. Therefore GSR data was employed as an index for stress. GSR data was checked throughout all trials and identified the behavior of subjects when GSR was increased (Fig. 5).

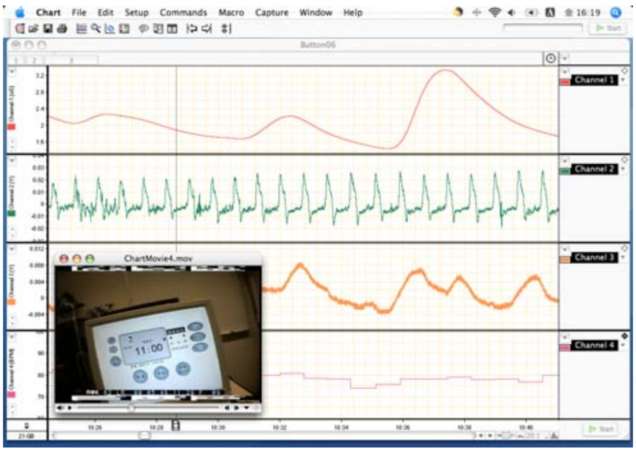


Fig. 5. Example of recorded physiological data and eye tracking data

First of all, increase of GSR was observed when the subjects considered which menu key should be pressed. Especially, while comparing "Rice cooking menu" and "Easy menu" of Stimulus B or "Menu", "Cooking type", and "Course" of Stimulus C increase of GSR was observed in three subjects each. As for Stimulus C, one subject wondered after pressing "Course" key for a while and GSR increased there.

If a key press had given different feedback than the subjects expected or no feedback, GSR has also increased. As for the selection of one menu key, the aimed cooking menu was not appeared by pressing the selected key in some cases. For instance, for the "Course" key of Stimulus C not rice cooking menu but other cooking menus such as "baking bread" were assigned and it confused the subjects. The keys for timer setting labeled with a triangle gave also different feedback to the subjects. One subject thought that the key "▼" in Stimulus A advanced the time. After pressing this key, this subject recognized that the time was set back. This feedback was unexpected for this subject. It was obvious in the comment "It works in another way than I expected."

Increase of GSR due to lack of feedback was typically observed at the beginning of the timer setting. In order to start timer setting, "Programming" key (or similar key) should be pressed. In most cases, the subjects pressed the keys to set the time first of all and then recognized that no feedback was given.

It was already pointed out that a key was sometimes pressed too many times so that they went past the aimed menu or time. After the subjects recognized it, GSR has increased. Also during repeated key press, increase of GSR was observed. Each key press switched menu or time, so that the subjects should pay attention to the display not to go past the aimed menu or time.

IV. DISCUSSIONS

A. Difficulties for older users

Behavior data, eye tracking data, and physiological data showed various difficulties for older users in use of modern rice cookers.

The first type of difficulties was related to the allocation of functions to operation keys and their labeling. Modern rice cookers can cook several types of rice in various ways and also other foods than rice. The problem was different allocation of these many different kinds of functions to operation keys. In this experiment, four stimuli provided three menu keys. In Stimulus A, "Rice" key was used to select type of rice, "Hardness" key to select texture of cooked rice (hard, medium, and soft), and "Menu" key to select rice cooking or other rice recipes (porridge and mixed rice). The function allocation in Stimulus B and C were similar to that of Stimulus A, but selection of type of rice and rice recipes was allocated together to one key ("Rice cooking menu" in Stimulus B and "Menu" in Stimulus C). "Easy menu" key in Stimulus B and "Course" key in Stimulus C was used to select other recipes than rice cooking. In Stimulus D, the selection of type of rice, texture of cooked rice, and rice recipes was altogether allocated to "Rice cooking menu" key. For pre-washed rice, an extra key was provided. "Healthy menu" key was applied for rice and non-rice recipes: to cook sprouted brown rice, tofu, yoghurt, and soft-boiled egg. Only Stimulus E provided single menu key. Type of rice, texture of cooked rice, and rice recipes could be selected by the key labeled "Course". Except Stimulus B and C, function allocation was different each other. In addition, the labeling was also different among stimuli. Especially similar labels such as "Menu" and "Course" were difficult to distinguish. The labels "Easy menu" and "Healthy menu" were hard to imagine which kind of menus was available by pressing these keys. Confusion of subjects elicited by different function allocation was clearly revealed by eye tracking data and GSR data. Gaze rate on single menu key in Stimulus E was obviously lower than total gaze rate on three menu keys in other stimuli. Increase of GSR was frequently observed during comparing menu keys. As results, some subjects pressed a wrong menu key.

Such difficulties are actually caused also by the fact that rice cookers offer too many functions. In interview, subjects said that they use usually only limited functions such as cooking white rice. Even timer cooking was not used by some subjects. It means that the tasks in this experiment were unfamiliar for the subjects. Without enough knowledge, the task accomplishment might be hard for the subjects.

The second type of difficulties was related to the interpretation of labels of keys for time setting. The time setting keys in Stimulus A and D were labeled by a triangle. However, the relationship between direction of the triangle and time transition produced by key press was different between two stimuli. In addition to such inconsistency, it matters also with interpretation. According to stereotype, an upward triangle "▲" means "up" and a downward triangle "▼" means "down". "Up" can be interpreted as "advance" and "down" as "reverse": then an upward triangle means "advance" and a downward triangle means "reverse". If time flow is concerned, however, it can be thought that time flows from top to down. In this case, an upward triangle should mean "reverse" in time flow and a downward triangle corresponds to "advance". Different from these labels, the labels "Hour" and "Minute" were clearly defined and easy to understand how they work.

The difference in key arrangement could also cause some difficulties, but it was not serious for older subjects. In the case of activation of timer in Stimulus B, the right upper part was frequently gazed, where "Rice cooking" key was provided in other four stimuli. Here the subjects seemed to get a hint from the consistent arrangement of the key to activate the timer. However, such behavior was not frequently observed and different key arrangement caused no confusion.

B. Proposal for a "better" user interface of home appliances

The results implicated that user interface for elderly users should provide enough information for function allocation. It should be realized from two aspects. First, allocation of menus or functions should be made logically. Each menu group should be clearly distinguished. Second, key labels should be explicit one: they should be easy to distinguish and to imagine how they work. Labeling is strongly related to the function allocation. If menus are categorized definitely, labeling should be easy.

If a user interface only for older adults is required, the needs of older users should be examined to identify the unnecessary menus. Simplification of the functions leads to simplification of the menu structure. If all the menus can be integrated into one key, the problem with function allocation will be solved.

V. CONCLUSION

In this experiment, the user interfaces of real rice cookers were examined by behavior observation, eye tracking, physiological measurement, and interview. Based on objective evaluation, it was revealed that the older subjects had difficulties when two or more similar operation keys were provided. Such difficulties were related with function allocation and key labeling. In order to avoid confusion for older users, definite classification of functions and explicit labeling are needed.

REFERENCES

[1] H.-W. Wahl, H. Mollenkopf, "Impact of everyday technology in the home environment on older adults' quality of life," in *Impact of technology on successful aging*, N. Charness, K. W. Schaie, Eds. New York: Springer Publishing Company, 2003, pp. 215-241.

[2] A. D. Fisk, W. A. Rogers, C. Neil, S. J. Czaja, J. Sharit, *Designing for older adults*. Boca Raton, FL: CRC Press, 2004, ch.2.

[3] R. Fukuda, H. Bubb, "Eye tracking study on Web-use: Comparison between younger and elderly users in case of search task with electronic timetable service," *PsychNology Journal*, vol. 1, pp. 202-228, 2003.

[4] R. D. Ward, P. H. Marsden, "Physiological responses to different WEB page designs", *International Journal of Human-Computer Studies*, vol. 59, pp. 199-212, 2003.

[5] T. Lin, A.Imamiya, M. Omata, W. Hu, "An empirical study of relationships between traditional usability indexes and physiological data," *Australasian Journal of Information Systems*, vol. 13, pp. 105-117, 2006.

[6] M. Thüring, S. Mahlke, "Usability, aesthetics and emotions in human-technology interaction", *International Journal of Psychology*, vol. 42, pp. 253-264, 2007.

[7] R. J. K. Jacob, K. S. Karn, "Eye tracking in human-computer interaction and usability research: Ready to deliver the promises," in *The Mind's Eye: Cognitive and Applied Aspects of Eye Movement Research*, J. Hyönä, R. Radach, H. Deubel, Eds. Oxford: Elsevier Science BV, 2003, pp. 573-605.