

Menu Selection Using Auditory Interface

Koichi Hirota¹, Yosuke Watanabe², and Yasushi Ikei²

¹ Graduate School of Frontier Sciences, University of Tokyo
5-1-5 Kashiwanoha, Kashiwa, Chiba 277-8563
{hirota,watanabe}@media.k.u-tokyo.ac.jp

² Faculty of System Design, Tokyo Metropolitan University
6-6 Asahigaoka, Hino, Tokyo 191-0065
ikei@tmit.ac.jp

Abstract. An approach to auditory interaction with wearable computer is investigated. Menu selection and keyboard input interfaces are experimentally implemented by integrating pointing interface using motion sensors with auditory localization system based on HRTF. Performance of users, or the efficiency of interaction, is evaluated through experiments using subjects. The average time for selecting a menu item was approximately 5-9 seconds depending on the geometric configuration of the menu, and average key input performance was approximately 6 seconds per a character. The result did not support our expectation that auditory localization of menu items will be a helpful cue for accurate pointing.

Keywords: auditory interface, menu selection, keyboard input.

1 Introduction

As computers become small and portable, requirement to use such computers all the time to assist users to perform their tasks from the aspect of information and communication. Concept of wearable computer presented a concrete vision of such computers and styles of using them^[1].

However, wearable computers still have not been commonly used in our life. One of the reasons is thought to be that the user interface is still not sophisticated enough for it to be used in daily life; for example, wearable key input device is not necessarily friendly for novice users and visual feedback through HMD is sometimes annoying while users' eyes are focusing on objects in real environment.

Some of these problems of the user interface are thought to be solved by introducing auditory interface where information is presented to the user through auditory sensation and interaction with the user is performed based on auditory feedback^[2]. Some experimental studies have been carried by ourselves^[3,4].

A merit of using auditory interface is that presentation of auditory information is possible simply using headphones. In recent years, many people are using headphones even while they are in public space, and that fact suggests that it can be used for long hours of wearing and listening. Also, headphones are not so weird as HMDs even if they are used in public space.

A drawback of auditory interface is that the amount of information presented through auditory sensation is generally much less than visual information provided by a HMD. This problem of auditory interface leads us to investigating an approach to improving the informational efficiency of the interface.

One of fundamental idea to solve the problem is an active control of auditory information. If the auditory information is provided passively, the user has to listen to all information that is provided by the system till the end even when the information is of no interest. On the other hand, if the user can select information, the user can skip items that are not required for the user, and it improves informational efficiency of the interface.

In the rest part of this paper, our first-step study on this topic is reported. Menu selection and keyboard input interfaces are experimentally implemented by integrating simple pointing interface with auditory localization, and their performance is evaluated.

2 Auditory Interface System

An auditory display system was implemented for our experiments. The system consists of an auditory localization device, two motion sensors, a headphone, and a notebook PC.

The auditory localization device is a dedicated convolution hardware that is capable of presenting and localizing 16 sound sources (14 from wave data and 2 from white and click noise generators) using HRTF^[5]. In the following experiments, HRTF data from KEMAR head^[6] was used. The motion sensors (MDP-A3U7, NEC-Tokin) were used to measure the orientation of user's hand and head. The sensor for head was attached to the overhead frame of the headphone, and the other sensor for hand was held by the user. Each sensor has two buttons whose status, as well as motion data, can be read by the PC. The notebook PC (CF-W4, Panasonic) controlled the entire system.

3 Menu Selection

The goal of this study is to clarify completion time and accuracy of menu selection operation. A menu system as shown in Figure 1 is supposed; menu items are located at even intervals of horizontal orientation, and user selects one of them through pointing it by hand motion sensor and pressing a button. The performance of operation was evaluated by measuring completion time and number of erroneous selections performed by the user under different conditions regarding number of menu items (4, 8, 12), angular width of each menu (10, 20, 30deg), with or without auditory localization, and auditory switching modes (direct and overlap); 36 in total combinations.

In case of without auditory localization, the sound source was located in front of the user. Auditory pointer means the feedback of pointer orientation by localized sound, and a repetitive click noise was used as sound source. The auditory switching mode means the way of switching auditory information when the pointer passes

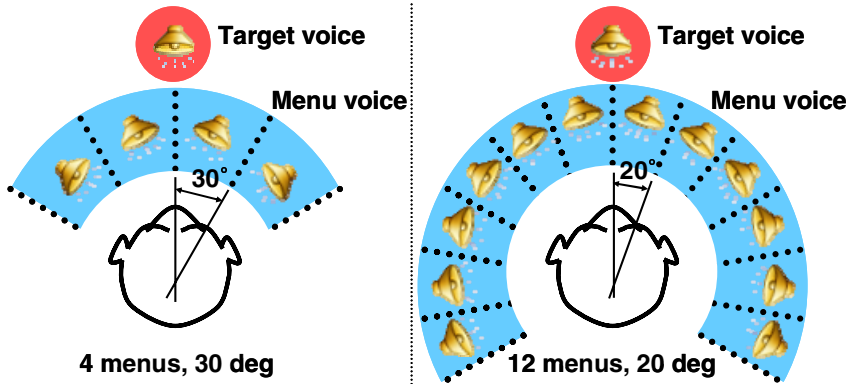


Fig. 1. Menu selection interface. Each menu items are arranged around the user at even angular intervals.

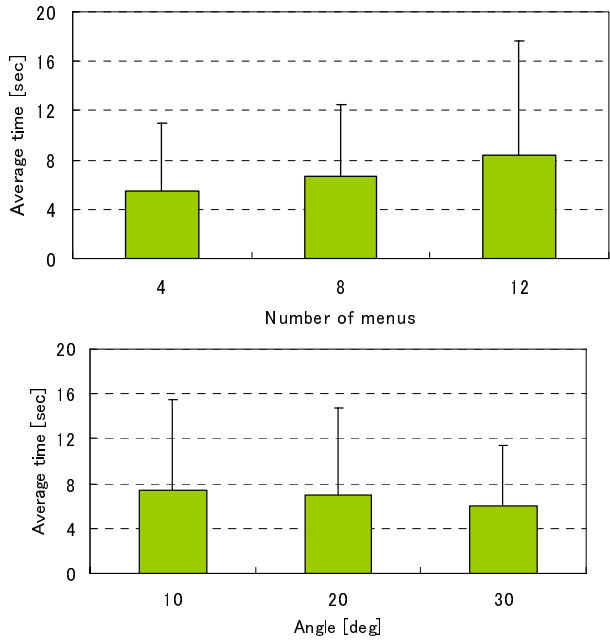


Fig. 2. Average completion time of menu selection. Both increase in the number of menu items and decrease in angular width of each menu item cause the selection task more difficult to perform.

across item borders; in direct mode, the sound source was immediately changed, while in overlap mode, the sound source of previous item continues existing until the end of pronunciation.

To eliminate semantic aspect of the task, vocal data from pronunciation of 'a' to 'z', instead of keywords of practical menus, were used for menu items. Volume of sound

was adjusted by the user for comfort. The sound data for menu items were randomly selected but without duplication. The number of subjects was 3, adult persons with normal aural ability. Each subject performed selection for 10 times for each of 36 conditions. The order of condition was randomized.

The average completion time computed for each conditions of the number of items and item angular width is shown in Figure 2. The result suggests that selection task is performed in about 5-9 seconds in average depending on these conditions. Increase in the number of items makes the task more difficult to perform and in both the difference among the average values was statistically significant ($p < 0.05$). Decrease in the size of item also makes the task difficult, and the difference was also significant ($p < 0.05$). On the other hand, the effect of both auditory localization and auditory presentation mode on the completion time was not made clear ($p > 0.05$).

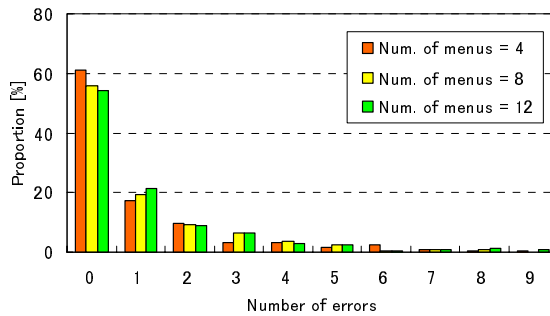


Fig. 3. Number of erroneous selections before selecting collect target item. Only one half of selection operation is performed successfully without retry.

The histogram of number of erroneous selection is plotted in Figure 3. The result suggests that only approximately 50% of selection is completed without error and retry. The success ratio is significantly low if it is compared with similar task using visual feedback. One reason of the result will be because the subjects were instructed to perform the task as fast as possible. The change in the number of menu items caused no noticeable difference in the histogram. Similarly, other conditions had no significant effect on the result.

4 Auditory Keyboard Input

As a more complicated case of menu selection interface, a keyboard input interface was experimentally implemented. Each key is considered as a menu item that is arranged in two-dimensional area as shown in Figure 4.

A map of qwerty keyboard was auditory presented in a similar way to auditory menu interface. In this interface the elevation angle of the pointer is also considered to allow two-dimensional selection. Other framework of the interaction is identical with the menu selection interface. The performance of the key input was measured regarding completion time and number of erroneous selection, under conditions of with and without auditory localization. The angular size of the map was fixed, and direct mode was used as auditory presentation mode.

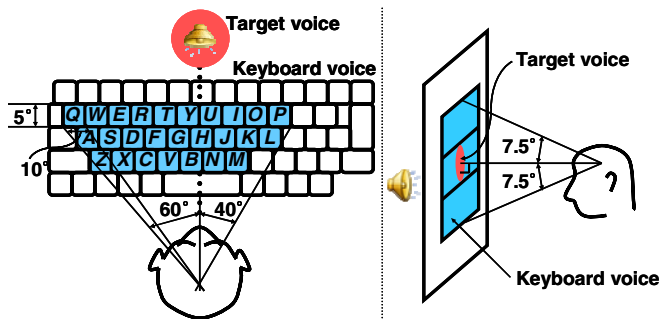


Fig. 4. Auditory keyboard interface. Two-dimensional keyboard layout is mapped to azimuth-elevation space in front of user.

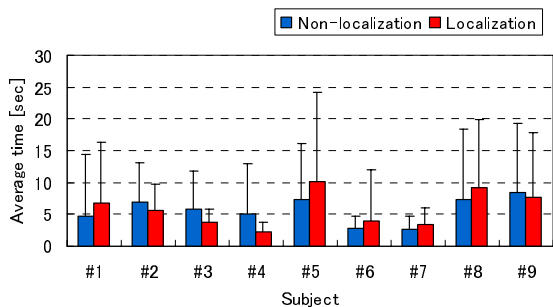


Fig. 5. Completion time of key input operation. There are large differences among individuals. There are some subjects who can better perform the task under the condition with auditory localization.

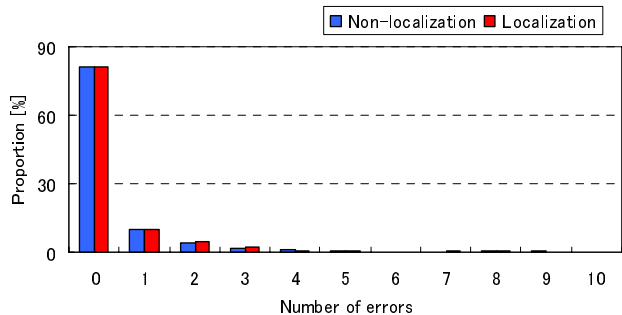


Fig. 6. Number of errors in key input operation. Error ratio of key input task is lower than the menu selection task, probably because the location of each item (or key) is predefined.

The number of subjects was 9, adult persons with normal aural ability. Each subject performed 40 input operations under each localization condition. The target

key was randomly chosen from alphabet 26 characters, and order of conditions was also randomized.

The average input completion time was approximately 6 seconds per a character irrelevant to auditory conditions; no significant difference in the completion time between two auditory conditions were found ($p>0.05$). A better performance compared with the menu selection interface is attained despite higher complexity of the task, because the arrangement of items (or keys) is familiar to the subjects. The individual difference of the completion time is shown in Figure 5. The difference may be caused by the difference about how each subject is used to qwerty keyboard.

The histogram about the number of errors is plotted in Figure 6. The result also suggests that more accurate operation is performed than the menu selection interface.

5 Conclusion

In this paper, an approach to auditory interaction with wearable computers was proposed. Menu selection and keyboard input interfaces were implemented and their performance was evaluated through experiments. The result did not support our expectation that auditory localization of menu items will be a helpful cue for accurate pointing. In our future work, we are going to investigate the users' performance in practical situation, such as while walking in the street. Also we are interested in analyzing the reason why auditory localization is not effectively used in the experiments that were reported in this paper.

References

1. Mann, S.: Wearable Computing. A first step toward Personal Imaging, IEEE Computer 30(3), 25–29 (1997)
2. Mynatt, E., Edwards, W.K.: Mapping GUIs to Auditory Interfaces. In: Proc. ACM UIST'92, pp. 61–70 (1992)
3. Ikei, S., Yamazaki, H., Hirota, K., Hirose, M.: vCocktail: Multiplexed-voice Menu Presentation Method for Wearable Computers. In: Proc. IEEE VR 2006, pp. 183–190 (2006)
4. Hirota, K., Hirose, M.: Auditory pointing for interaction with wearable systems. In: Proc. HCI 2003, vol. 3, pp. 744–748 (2003)
5. Wenzel, E.M., Stone, P.K., Fisher, S.S., Foster, S.H.: A System for Three-Dimensional Acoustic 'Visualization' in a Virtual Environment Workstation. In: Proc. Visualization '90, pp. 329–337 (1990)
6. Gardner, W.G., Martin, K.D.: HRTF measurements of a KEMAR dummy head microphone. MIT Media Lab Perceptual Computing Technical Report #280 (1994)