Gaze-Directed Hands-Free Interface for Mobile Interaction

Gie-seo Park, Jong-gil Ahn, and Gerard J. Kim

Digital Experience Laboratory
Korea University, Seoul, Korea
{2002,hide989,gjkim}@korea.ac.kr

Abstract. While mobile devices have allowed people to carry out various computing and communication tasks everywhere, it has generally lacked the support for task execution while the user is in motion. This is because the interaction schemes of most mobile applications are centered around the device visual display and when in motion (with the important body parts, such as the head and hands, moving), it is difficult for the user to recognize the visual output on the small hand-carried device display and respond to make the timely and proper input. In this paper, we propose an interface which allows the user to interact with the mobile devices during motion without having to look at it or use one's hands. More specifically, the user interacts, by gaze and head motion gestures, with an invisible virtual interface panel with the help of a head-worn gyro sensor and aural feedback. Since the menu is one of the most prevailing methods of interaction, we investigate and focus on the various forms of menu presentation such as the layout and the number of comfortably selectable menu items. With head motion, it turns out 4x2 or 3x3 grid menu is more effective. The results of this study can be further extended for developing a more sophisticated non-visual oriented mobile interface.

Keywords: Mobile interface, Gaze, Head-controlled, Hands-free, Non-visual interface.

1 Introduction

Mobile interaction has become an important issue with the explosive usage of hand-held devices. Most mobile interfaces are suited only for the situation where the user is stationary and interacting closely (visually) and stably with the device in hand. However, there are also many occasions where one needs to interact with the mobile device while in motion. In this situation, it is often difficult to make out (let alone use) the visual display and directly interact with the objects (on the touch screen or hardware buttons) effectively.

On the other hand, sensors external to the hand-held mobile device, but integrated with embedded in the body accessories (e.g. eye glasses, headset, ear piece, hat, necklace) are becoming popular and opening new opportunities in mobile interaction area.

J.A. Jacko (Ed.): Human-Computer Interaction, Part II, HCII 2011, LNCS 6762, pp. 304–313, 2011.

In this paper, we propose an interface which allows the user to interact with the mobile devices during motion without having to look at it (non-visual) or use one's hands (hands-free). Being non-visual and hands-free, the user is allowed to carry on the main motion task without visual distraction. Specifically, our interface proposal is to use gaze/head motion gestures, with an invisible virtual interface panel with the help of a head-worn gyro sensor and aural feedback. Furthermore, as the menu is one of the most prevailing methods of interaction, we investigate and focus on the various forms of menu presentation, for the proposed interface, such as the most effective and usable layout and the number of comfortably selectable menu items.

This paper is organized as follows: a brief description of related research is given in the next section. Section 3 provides the implementation details of the proposed interface. The menu design space exploration experiment is presented in Section 4 with the results discussed in Section 5. Finally we conclude the paper with an executive summarization of our contribution and directions for future work.

2 Related Work

Recent smart phones are equipped with many sensors such as the GPS, tilt sensor, accelerometer, digital compass, camera, etc. In relation to our work, many interfaces taking advantage of these sensors have appeared, e.g. for pointing and scrolling [14], menu marking [13] changing screen orientation [8], zooming and navigation [5] and shaking based event enactment [17]. However, most of these attempts were still centered on the mobile device itself and require visual recognition of the interaction mode and results.

For eyes-free operation, Oakley and O'Modhrain [12] proposed a tile based interaction with tactile feedback for menu navigation. Brewster et al. also developed an eyes-free interface that used 3D aural cues for choosing menu items [2]. While eyes-free, note that these interfaces still require the interaction to be applied in a stable non-moving condition.

With hands occupied, the human head can act as a helpful input medium, providing directional or gestural cues. Head/gaze based interaction has been applied with mixed results for non-mobile computing environments [1][18], but to a less extent for mobile situations. Crossan et al. [3] investigated the possibility to use head tilting for mobile interaction during motion. While their interface still required the use of hands and eyes, our work was inspired by their approach.

3 Interface Design

3.1 Basic Idea: Gaze and Sound Directed

The basic idea is to build a head-controlled and hands-free interface so as to select a menu item while a user is in motion and hands occupied (walking, running or driving). As an application is started, the existence of a virtual pop-up menu is notified to the user via voice and enters the menu selection mode. The user is instructed to first center one's head. Then, a virtual ray emanating from the user's head, in the direction of one's gaze, to a virtual menu panel/item situated in the front.

Since the menu is virtual, interaction with the menu system is guided through sound and voice (e.g. which menu items are where). The gaze is used to select a menu item and the final confirmation is given by shaking/nodding one's head or by gazing for a fixed amount of time, also accompanied with a final aural feedback (Figure 1).

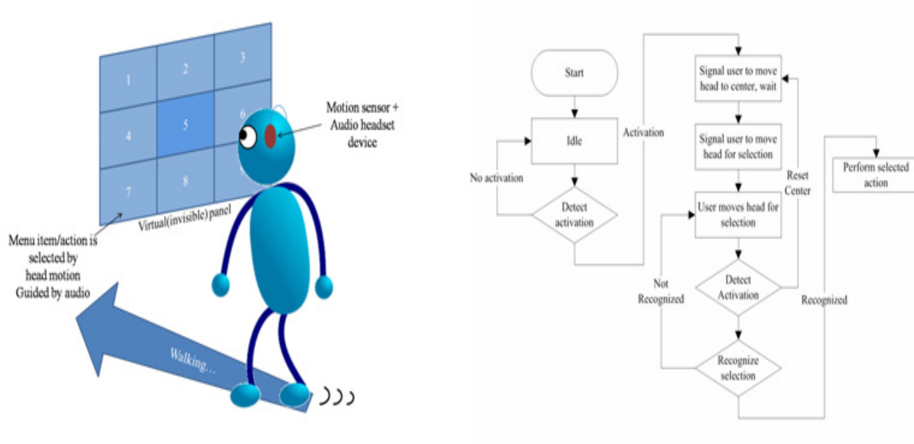

Fig. 1. The proposed gaze and sound directed menu interface

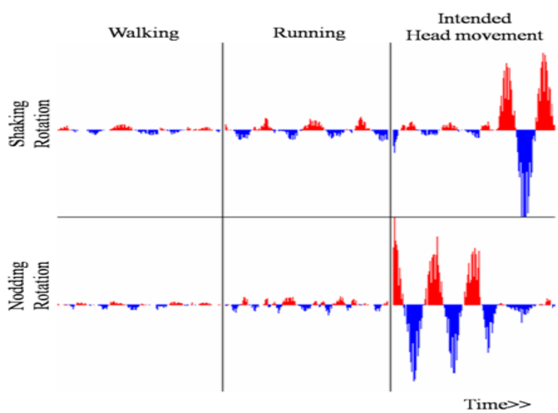

Fig. 2. Sensor values of the head-attached gyrometer for intentional nodding or shaking over walking or running motions

The head movement and direction is detected by a wirelessly operated (Bluetooth) gyro sensor worn on the head (note that even if the sensor is not attached to the main mobile device, such a configuration is deemed quite acceptable and not so inconvenient from the example of already popular Bluetooth operated headsets). A gyro sensor was used (e.g. over an accelerometer) due to the easier discerning between true head motion and mere body motion. Figure 2 shows that indeed the intended nod or shaking was easily detected over the running or walking motion. In actual implementation for the experiment, we attached the G2 Wireless Mouse (from Ocean Technology, which uses the MG1101 dual axis gyrometer form the Gyration Corporation) to a cap (see Figure 3).

Fig. 3. Wireless gyrometer sensor attached to a cap

The virtually selectable menu panels can vary in several aspects, e.g. in its form, the use of gestural selection method (as an alternative to gaze based selection, Figure 4), confirmation method (Figure 5), etc. Note that the selection and confirmation methods are coupled with the form of the menu. For instance, with 2D menus, selection can be made by moving over the panel in “4-ways” and making a “Round trip” gaze movement for confirmation. Such a method would not be possible with a 1D menu.

The numbers of menu items is set by equally dividing user’s effective head angle range. Human head normally can be rotated about 140 degrees for shaking and 100 degrees for nodding [9]. However, rotating the head in such a wide range is practically not feasible, since the user needs to attend to the other on-going task.

For this reason, calibration process for a user to measure the neck rotating range was performed. The range was detected while a user rotates head with keeping the vision for moving activity.

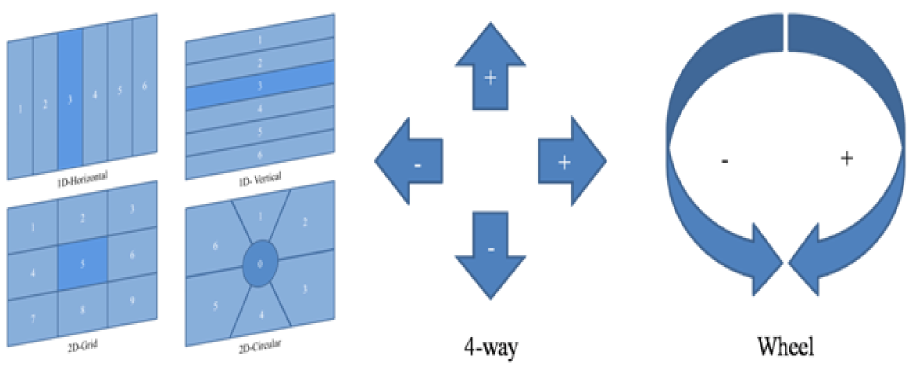

Fig. 4. Various forms of virtual menu and examples of gestural methods for menu selection (as an alternative to gaze based selection)

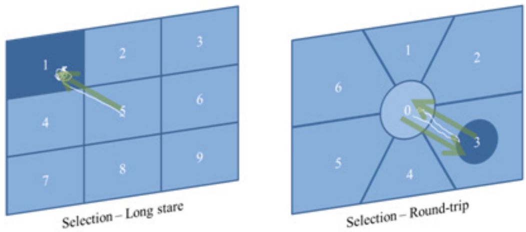

Fig. 5. Two examples of confirmation method: "Long stare" vs. "Round trip"

4 Experiment

A usability experiment was carried out to find the preferred form of the virtual menu and the proper number of effectively selectable menu items (in a fixed area).

4.1 Experiment Design

The two factors in the experiment was the type or form of the menu and the number of menu items. This resulted in about total of 19 treatment groups. For all treatments, gaze based selection was used and both "Long-stare" and "Round-trip" based confirmation methods were tested (see Figure 5).

Table 1. 19 tested combinations between the type of menu and no. of menu items

Menu Form	No. of Menu Items
1D Horizontal	2, 3, 4, 5, 6, 7, 8
1D Vertical	2, 3, 4, 5, 6, 7, 8
2D Grid	4 (2x2), 6 (3x2), 8 (4x2), 9 (3x3), 4x3 (12)

The user was given a menu selection task and the total accumulated selection time was used as the main dependent variable reflecting of the user performance. Selection error was also recorded. To help precise selection, the selected item was enlarged (see Figure 6). Even though the enlarged (twice the size of the menu item) virtual menu item is not visible, the enlarged size helps the user stay with the menu item within a tolerable range and makes the interaction robust to noise and nominal user shaking.

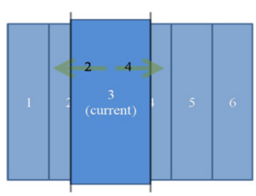

Fig. 6. Increasing the tolerance for a selected menu item (for reducing error and making the system robust to sensor/motion noise)

4.2 Experiment Process

The experiment proceeded in two stages: first with the 1D menus to find the upper limit for the number of menu items for 2D menus in each horizontal and vertical dimension. 1D horizontal and vertical menus with 2 to 8 items (smaller relative menu item width with increasing number of items) were tested first. The experiment was conducted in motion on a tread mill with the speed set at 4km/hour.

Based on this experiment, the upper limit for the 2D menu items were set at 4hx3v. 2D menus with 4x3, 4x2, 3x3, 3x2 and 2x2 configurations were tested in the second stage of the experiment. Each experiment was run as a two factor within-subject repeated measure. During the experiments, the subject was given a particular menu configuration in a balanced order and asked to carry out a menu selection task, three times. The first trials and extreme outliers were excluded from the analysis. The subject's task performance was recorded in terms of the accuracy and time of completion. A general usability survey was also taken afterwards.

5 Experiment Results

5.1 1D Menu

The experiment results (proper number of menu items) for the 1D (horizontal and vertical) menus are shown in Table 2 and Figure 7. The results show that as the menu item increases (and thus the relative size of the menu items decrease), the selection time increases and the accuracy is reduced (obviously). However, in general, the horizontal 1D menu is faster and more accurate, despite the slightly longer operation range (than that of the vertical 1D menu given the same number of menu items/divisions). The 1D horizontal menu is also more stable (less deviation).

Table 2. The selection time and accuracy for 1D horizontal and vertical menus

Division		8	7	6	5	4	3	2
Horizontal (L/R shake)	Accuracy (%)	66.7	80.0	100.0	100.0	100.0	100.0	100.0
	Average (sec)	5.8	5.2	4.6	3.6	3.2	3.1	2.6
	StD (sec)	0.6	0.2	0.2	0.2	0.1	0.2	0.03
Vertical (U/D nod)	Accuracy (%)	0.0	66.7	72.7	88.9	100.0	100.0	100.0
	Average (sec)		6.1	5.5	4.8	4.0	2.7	2.6
	StD (sec)		1.5	1.1	0.8	0.6	0.06	0.06

5.2 2D Menu

Based on the 1D result, the maximum grid division for the 2D menus was decided to be 4x3 (e.g. average selection time < 4 seconds and standard deviation < 0.2 seconds). Starting from 4x3, we reduced the 2D menu configuration down to 2x2 in the experiment.

For 2D menus, 4x2 and 3x3 grid menus proved to be the optimal configurations (in consideration of the selection time and associated accuracy). Interestingly, the number of menu items (8-9) coincided with the capacity of the human short term memory [11]. Table 3 shows the experimental results (selection accuracy and average task completion time/std.).

Table 3. The selection time and accuracy for 2D menus

Division	4x3	4x2	3x3	3x2	2x2
Accuracy (%)	88,9	100.0	100.0	100.0	100.0
Average (sec)	4.9	3.4	3.4	3.5	3.2
StD (sec)	0.7	0.2	0.2	0.2	0.3

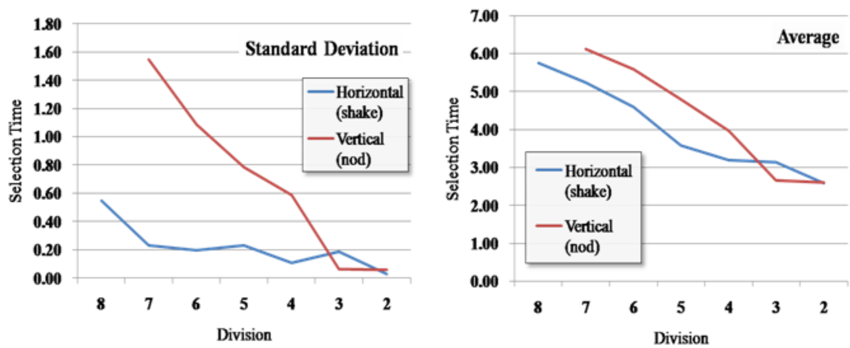

Fig. 7. 1D menu results for selection time (standard deviation and average)

5.3 Circular Menus

In a separate/follow-up experiment, 2D circular menus were also tested. For the same number of menu items, compared to grid type, the circular menus did not perform well especially accuracy-wise. This was due to the slight confusion with the number layout being different from the clock. Also the "Round trip" style of menu item confirmation resulted in slower performance and high error (in having to deal with two spatial locations).

5.4 ANOVA Result

Analysis of variance (ANOVA) was applied to the experimental data (total selection time) and the main effect of the different menu forms/sizes was verified (p-value < 0.000 and F-value = 24.266) as well. The post-hoc test also showed such a similar trend and grouping among the menu types/sizes (we omit the detailed statistic values).

Table 4. Analysis of variance (total selection time)

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	244.29	21	11.63	24.27	.000
Within Groups	80.54	168	.479		
Total	324.83	189			

5.5 Fitt's Law

The resulting performance may be explained by the Fitt's law [6] even though the proposed interface uses gaze and head movement, instead of the hand as was [15], to move the cursor/object and accomplish the given task. According to the Fitts' Law and other variations [10][16] the movement time (task time, MT) to select a target of width W at the distance D is related as follows:

$$MT = a + b\log_2(\frac{D}{W} + c)$$

where a, b and c are constants induced by a linear regression. The value inside the log function is also called the index of difficulty (ID), and ID is to exhibit a linear relationship with the MT. The experimental data, as shown in Figure 8, generally adhered to the Fitt's Law.

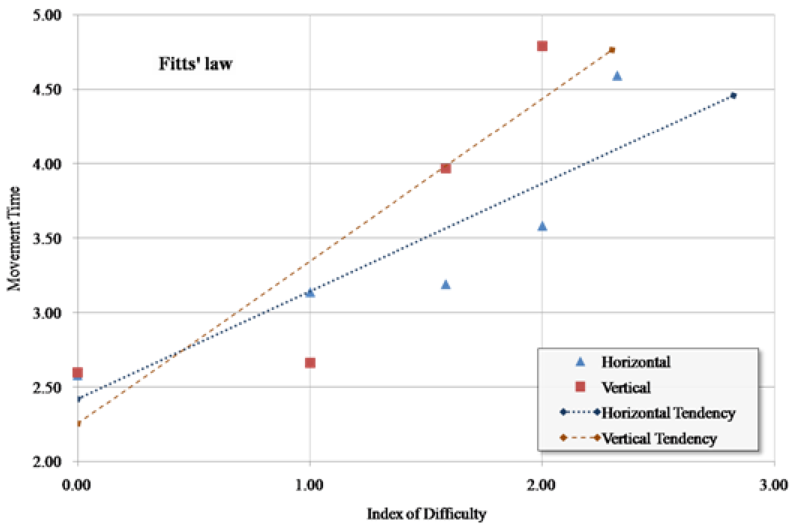

Fig. 8. The average movement time (MT) plotted against Fitts' index of difficulty (ID)

6 Conclusion

Interaction during motion will become a very important issue in mobile interfaces. Increasing number of people are interacting during motion, e.g. watching TV/video, talking on the phone, playing games. Provisions are needed to ensure people can carry out these multiple tasks concurrently e.g. by leveraging on different modalities.

In this paper, we presented a way to interact with a mobile device without looking at it and using hands that is by head motion only, in motion. With our proposed interface using the head motion and gaze, it turns out 4x2 or 3x3 grid menu is the most effective configuration (near 100% recognition, low std. and relatively more menu items) considering the number available menu items vs. selection time and associated error rate. Between the horizontal and vertical, the horizontal head movement was more stable with longer operational range (and thus more menu items covered).

In particular, non-visual interface can also be used for visually impaired users. In addition the findings can be applied to designing augmented reality interfaces that uses head motion or gaze. This study is only the first step in the investigation of non-visual mobile interfaces. Also equally required is more stable sensing for filtering out true intentional head/gaze motion from normal moving activities (e.g. navigation) and precise virtual cursor tracking.

Acknowledgement. This research was supported in part by the Strategic Technology Lab. Program (Multimodal Entertainment Platform area) and the Core Industrial Tech. Development Program (Digital Textile based Around Body Computing area) of the Korea Ministry of Knowledge Economy (MKE).

References

1. F. Berard, “The Perceptual Window: Head Motion as a new Input Stream”, IFIP Conference on Human-Computer Interaction, 238–244 (1999).
2. S. Brewster, J. Lumsden, M. Bell, M. Hall, and S. Tasker, “Multimodal 'Eyes-Free' Interaction Techniques for Wearable Devices”, Proc. of ACM CHI, Florida, 463-480 (2003).
3. Crossan, M. McGill, S. Brewster and R. Murray-Smith, “Head Tilting for Interaction in Mobile Contexts”, Proc. of the 11th Intl. Conf. on Human-Computer Interaction with Mobile Devices and Services, MobileHCI, Bonn, Germany (2008).
4. Crossan, J. Williamson, S. Brewster, and R. Murray-Smith, "Wrist Rotation for Interaction in Mobile Contexts", Proc. of Mobile HCI, Amsterdam, Netherlands, (2008).
5. P. Eslambolchilar, and R. Murray-Smith, “Control Centric Approach in Designing Scrolling and Zooming User Interfaces”, International Journal of Human-Computer Studies, 66(12), 838-856 (2008).
6. P. Fitts, “The Information Capacity of the Human Motor System in Controlling the Amplitude of Movement”, Journal of Experimental Psychology, 47, 381-391 (1954).
7. E. Foxlin, “Inertial Head-Tracker Sensor Fusion by a Complementary Separate-Bias Kalman Filter”, Proc. of IEEE Virtual Reality Annual International Symposium, 184-196 (1996).
8. K. Hinckley, J. Pierce, and E. Horvitz, “Sensing Techniques for Mobile Interaction”, Proc. of UIST, 91-100 (2000).
9. E. LoPresti, D. Brienza, J. Angelo, L. Gilbertson, and J. Sakai, “Neck Range of Motion and Use of Computer Head Controls”, Proc. Intl. ACM Conference on Assistive Technologies, Arlington, Virginia, 121-128 (2000).
10. MacKenzie, “A Note on the Information-Theoretic Basis for Fitts' Law”, Journal of Motor Behavior, 21, 323-330 (1989).
11. G. Miller, “The Magical Number Seven, Plus or Minus Two: Some Limits on Our Capacity for Processing Information,” Psychological Review, 63(2), 343-355 (1956).
12. Oakley and M. O'Modhrain, “Tilt to Scroll: Evaluating a Motion based Vibrotactile Mobile Interface”, Proc. of World Haptics, 40-49 (2005).
13. Oakley and J. Park, “A Motion-based Marking Menu System”, Extended Abstracts of ACM CHI, San Jose (2007).

14. Rekimoto, “Tilting Operations for Small Screen Interfaces”, Proc. of UIST, 167-168 (1996).
15. G. Robert, C. Gregg, “A Method for Evaluating Head-Controlled Computer Input Devices Using Fitts’ Law”, Human Factors, 32(4), 423-438 (1990).
16. Welford, “Fundamentals of Skill”, London: Methuen, (1968).
17. Williamson, R. Murray-Smith, and S. Hughes, “Shoogle: Multimodal Excitatory Interaction on Mobile Devices”, Proc. of ACM CHI, San Jose (2007).
18. S. You and U. Neumann, “Fusion of Vision and Gyro Tracking for Robust Augmented Reality Registration”, Proc. of IEEE Conference on Virtual Reality, Japan, 71-78 (2001).