Effect of Icon Amount and Visual Density on Usability of Smartwatches

Fan Mo(✉), Shuping Yi, and Jia Zhou

Department of Industrial Engineering, Chongqing University, Chongqing 400044, China
mofan1992@gmail.com, ysp@cqu.edu.cn, zhoujia07@gmail.com

Abstract. Appropriate design for smartwatches menu on the tiny touchscreen is desired and, hence, this study aims to explore the effect of icon amount and visual density on older adults' performance and satisfaction. To achieve this goal, an experiment was conducted. A total of 15 older adults participated in this study. The results of this study showed which combination of visual density and icon amount was better on the tiny screen (e.g., 40 mm wide and 40 mm high). First, a small number of icon amount (e.g., 8 icons) was generally better used in smartwatches menu. In that case, both visual density of 4 icons per page and 8 icons per page caused better performance. Second, compared with 1 icon and 8 icons per page, the visual density of 4 icons per page might was the best choice if the number of icon amount was large (e.g., 24 icons and 48 icons). Finally, the study also indicated that older adults' gestures of wearing smartwatches on the back of wrist, taping with the index finger and closing hand related to better performance.

Keywords: Smartwatches menu · Icon amount · Visual density · Older adults · Users' gestures

1 Introduction

Smartwatches are in dilemma about how to display abundant applications on a tiny display with the increasingly used by older adults. They have the potential to seamlessly integrate with human body and integrate functions of health care and personal safety, which could benefit older adults in particular. But displaying functions on a tiny display is a great challenge for older adults. To solve this problem, the menu is widely used in smartwatches to organize applications.

The icon amount and visual density of menus might influence older adults' performance and satisfaction. Icon amount is the total number of applications in the menu, and visual density is the number of applications from per page. A large number of icon amount could result in long and complex menus, and a large number of visual density could increase older adults' cognitive load. Previous research found that proper smartphones menu' icon amount and visual density could improve interface usability. However, there is a lack of study on the icon amount and visual density of smartwatches. Therefore, this paper proposed different icon amount and visual density of smartwatches, developed prototypes, and tested their effect on older adults' performance and satisfaction through experiment.

J. Zhou and G. Salvendy (Eds.): ITAP 2016, Part I, LNCS 9754, pp. 466–477, 2016.
DOI: 10.1007/978-3-319-39943-0_45

2 Literature Review

In previous research, the menus of smartphones and smartwatches were studied. The limited display size could increase users memory load [1]. In order to reduce the memory burden on the users, the depth, breadth of menus should appropriate [2, 3]. There could be conflict with between menu breadth and font size. A study on feature phones showed that displaying five menu items per page was better than displaying one menu item per page, and 12pt font size was better than 8pt [4]. As to the smartwatch, four most common menus (one-line list, two-line grid menu, rotating menu, and clustered menu) were tested. The results showed that two-line grid menu and clustered menu were more efficient at performance [5].

As to the interaction of smartwatches, most of studies focused on the input of smartwatches. Existing studies can be classified as two branches: interaction outside the display or interaction inside the display. As to interaction outside the display, Perrault, Lecolinet, Eagan and Guiard [6] proposed a novel gesture technique for wristband interaction, in order to avoid the visual occlusion and the fat finger problem. And then, Knibbe et al. [7] extended the interactive surface for a smartwatch to the back of the hand and defined a range of supported bimanual gestures. As to interaction inside the display, text entry was redesigned to adapt to the tiny input screen and obtained a better result than the original interface [8, 9]. However, when users taped on the screen, the error rates became higher as the buttons became smaller [10]. In addition, Shen, Xue, Li and Zhou [11] indicated that the number of the icons should not be more than 25 in one area, and for small number of the icons, the interelement spacing should be more than 1/2 icon.

The study of the output of smartwatches is needed. This research chose icon amount and visual density as two factors, to find out their effect on performance and satisfaction of smartwatches.

3 Methodology

3.1 Variables

The independent variables were icon amount and visual density. Both of them were within-subject variables. Three levels of icon amount were 8 icons, 24 icons, and 48 icons. Three levels of visual density were 1 icon per page, 4 icons per page, and 8 icons per page.

Recent investigations indicated that there were average 95 installed apps on smartphones around the world [12], and there were average 36 installed third-party apps on smartphones in China [13]. Further, the existing smartwatches (Moto 360, Apple Watch, and Samsung Gear) showed that the available number of icon amount ranged from 6 to 64, and visual density of 1, 2 and 9 icons per page were appropriate choices. To ensure that visual density was appropriate and would not change in every page, icon amount of 8, 24 and 48 icons and visual density of 1, 4 and 8 icons per page were chosen.

The stimulus displayed nine prototypes on a smartwatch, and prototypes of three levels of visual density are shown in Fig. 1. The icon's dimensions were 6.30 mm (60px) wide and 6.30 mm (60px) high. The distance between two icons was 3.78 mm (36px) in the prototypes with 4 icons per page, and was 1.58 mm (15px) in the prototypes with 8 icons per page.

Fig. 1. The stimulus of three levels of visual density on Moto 360

The dependent variables were task effectiveness, efficiency and users' satisfaction. In total, six different dependent variables were surveyed. Two measures referred to effectiveness, three of the measures addressed efficiency, and one measure was related to users' satisfaction of using prototypes. The measures are described in detail below.

Effectiveness can be measured by the success rates of solving tasks, and the success rates of taping. Efficiency can be measured by the number of necessary swiping left in the menu, the number of swiping right, and task completion time. Every task started with the first page, so users needed to keep swiping left to complete the tasks. The measure of number of swiping right indicated that users in the belief of having crossed the correct position went back to a previous position within the menu.

Users' satisfaction was measured by questionnaire (PSSUQ). It has seven items. Six items measures user's perceived usability [14], and the last item is an overall question to measure usage intention [15]. Five-point Likert scale was used.

Four demographic variables were taken into account. User' basic information (age, education, etc.) was measured by questionnaire. User's finger dexterity and vision were measured by finger dexterity test and near vision test. Moreover, users' gestures of using the smartwatch were recorded.

3.2 Participants

Sixteen older adults (9 males and 6 females) took part in this experiment. Their average age was 65.47 (SD = 5.29, range 59 to 74). They were recruited from Chongqing Jiangbei district Yuzui aged community. Older adults who had poor eyesight were encouraged to wear presbyopia glasses. 73.3 % of older adults had Junior High School diploma. 26.67 % of them had used smartphones for 2 years. 33.33 % of them used to wear watches until they started to use mobile phones. Participants were highly interested in their performance and enjoyed this experiment very well.

3.3 Equipment

One pilot test was conducted in older adult' home. Formal experiment was conducted in an office of the Citizen School of Yuzui, Jiangbei, Chongqing. Prior to the task, a Finger Dexterity Tester (Beida Jade Bird, 2-601) and a Near Vision Test Card were used to measure the participants' finger dexterity and vision. In the experiment, a smartwatch (Moto 360; screen: 1.56″, 320 × 290, round; OS: Android Wear OS 5.1.1) and a smartphone (Xiaomi 2; screen: 4.3″, 1280 × 720; OS: Android OS 5.0.2, MIUI Global 7.1) were used. Moto 360 connected to Xiaomi 2 via Bluetooth, and the nine prototypes were running on them. Prototypes were developed by Java language. The coordinates and times of every tap and swipe were listened and recorded.

3.4 Experimental Tasks

These prototypes were tested in a random order. Participants were required to complete 45 tasks, and five tasks for each prototype. Each participant was requested to search for a target number in each task. The numbers were in a random arrangement.

In order to know the target number they were going to search in a task, participants needed to look at the smartphone screen at first (shown in Fig. 2). They searched for the target number by swiping left and right within prototypes and taped the icon with the correct number. Then, the number on the smartphone screen changed into the next target number of tasks. When a tap was detected, the background color of icon turned green to alert the participants. Meanwhile, the smartphone played the sound of a beep in order to draw participants' attention to the change of target number.

Fig. 2. Illustration of the display conditions in the smartphone and smartwatch

These prototypes were designed for simulating the scenario of using real smartwatches menu, so each task started with the first page. To record the start time in a prototype, each prototype contained a dialog box that was displayed before all tasks began, and participants taped to start the first task. After participants completed the tasks of one prototype, a dialog box of "End and Congratulations" displayed.

3.5 Procedures

The experiment took each participant about 60 min. The experiment process was listed as follows. First, participants were given an overall introduction of the experiment and were required to complete a questionnaire in order to provide their personal details. Under the Finger Dexterity Tester (shown in Fig. 3), participants were required to align steel pins at holes, and put them in the holes one by one from "Start" to "End". Participants' finger dexterity was measured through the total time taken to complete. Further, participants' vision was measured. Second, participants were shown a basic demonstration of the system. They practiced about the basic operation of prototypes and completed the tasks in this tutorial demonstrate. And then, participants were required to complete the 45 tasks. The details of user action in every task were recorded. Each participant was encouraged to complete the experiment individually, and was not allowed to talk with experimenters. After completing each prototype, participants filled in the PSSUQ. In the final of the experiment, participants were interviewed briefly. Two experiment field scenes are shown in Fig. 3.

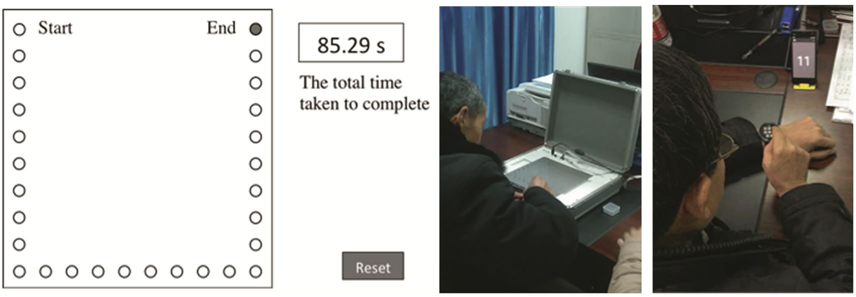

Fig. 3. Instruction of Finger Dexterity Tester and experiment field scenes (finger dexterity test and fulfilling tasks in the smartwatch)

4 Results and Discussion

4.1 Overall Performance

The influences of icon amount and visual density on effectiveness and efficiency were tested through MANOVA. The results showed that they did not significantly affect effectiveness. Regarding efficiency, there were significant differences in the number of swiping left ($F(2,126) = 100.3$, $p < 0.001$), the number of swiping right ($F(2,126) = 10.3$, $p < 0.001$), and task completion time ($F(2,126) = 88.2$, $p < 0.001$) among three levels of icon amount. What's more, there were significant differences in the number of swiping left ($F(2,126) = 168.2$, $p < 0.001$), the number of swiping right ($F(2,126) = 11.6$, $p < 0.001$), and task completion time ($F(2,126) = 55.4$, $p < 0.001$) among three levels of visual density. Therefore, multiple comparison was conducted.

As shown in Fig. 4, there were significant differences between icon amount of 8 and 24 icons ($t < 0.001$), 8 and 48 icons ($t < 0.001$), 24 and 48 icons ($t < 0.001$), and there were significant differences between visual density of 1 and 4 icons per page (t < 0.001), 1 and 8 icons per page (t < 0.001), 4 and 8 icons per page (t = 0.019) in the number of swiping left. Prototypes with larger number of icon amount or lower visual density had larger number of swiping left than the opposite prototypes. One possible reason was that breadth of the menu was increased, so participants needed to search the page one by one within the menu through swiping left.

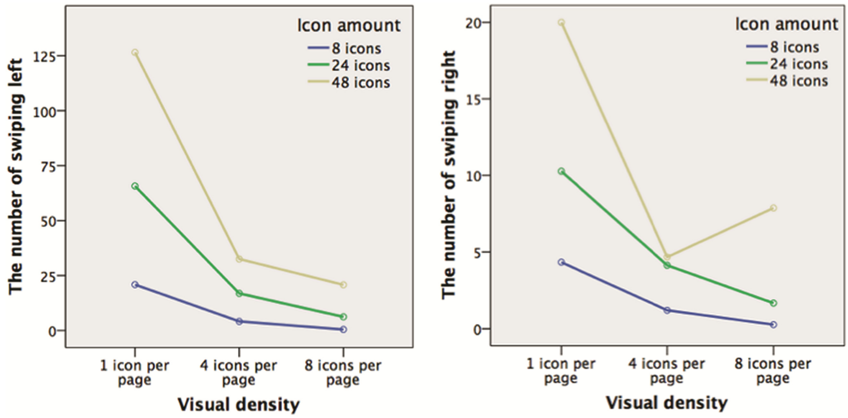

Fig. 4. The number of swiping left and the number of swiping right in three levels of icon amount and three levels of visual density

There were significant differences between icon amount of 8 and 48 icons (t < 0.001), 24 and 48 icons (t = 0.006), and between visual density of 1 and 4 icons per page (t < 0.001), 1 and 8 icons per page (t < 0.001) in the number of swiping right. Users went back to a previous position within the menu by swiping right, when they in the belief of having crossed the correct position. Which meant that icon amount of 48 icons or visual density of 1 icon per page led to that users were more likely to miss the target page.

As shown in Fig. 5, there were significant differences in the task completion time between icon amount of 8 and 24 icons (t < 0.001), 8 and 48 icons (t < 0.001), 24 and 48 icons (t < 0.001). That is, participants spent more time to search the target page with the increasing numbers of icon amount. There were significant differences in the task completion time between visual density of 1 and 4 icons per page (t < 0.001), 1 and 8 icons per page (t < 0.001). Similarly, low visual density led to the increasing of breadth of the menu, so participants spent more time when searching the target page. However, 4 icons per page had no significant differences with 8 icons per page. Which meant that searching in visual density of 4 and 8 icons per page were similar in task completion time for older adults.

The results indicated that icon amount and visual density has apparent influence on some aspects of efficiency. Further, multiple comparison was conducted to analysis the influence on effectiveness.

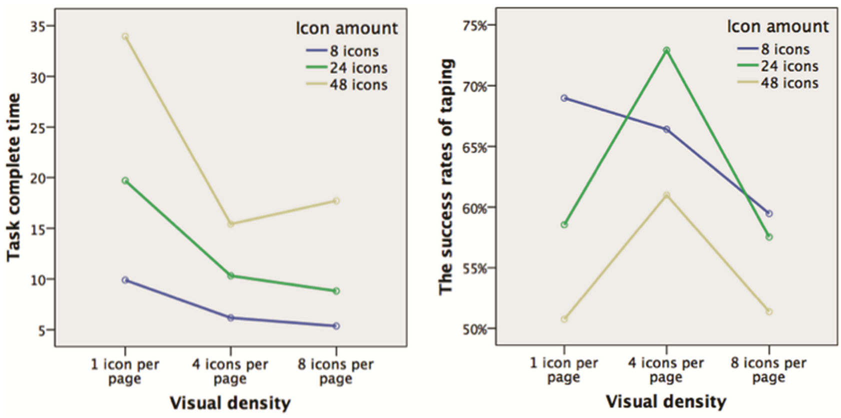

Fig. 5. Task completion time and the success rates of taping in three levels of icon amount and three levels of visual density

There were significant differences in the success rates of taping between icon amount of 8 and 48 icons (t = 0.025), and between visual density of 4 and 8 icons per page (t = 0.024). Icons amount of 48 icons had lower success rates of taping than 8 icons. One possible reason was that older adults carelessly taped the screen when they failed to swipe. That's one of the reasons why visual density of 1 icon per page had higher success rates of taping with the icon amount of 8 icons, but had lower success rates with 24 and 48 icons. Moreover, visual density of 8 icons per page was likely to lead to visual occlusion and the fat finger problem, which may further resulted in low success rates. And visual density of 4 icons per page had the highest success rates of taping compared with 1 and 8 icons per page. However, there were no significant differences in the success rates of solving tasks.

4.2 Users' Judgments with Respect to the Ease of Use

The average score of three levels of icon amount and three levels of visual density are shown in Fig. 6. The results indicated that participants preferred a menu with less number of icon amount or high visual density.

To dig into the reasons of the difference of users' satisfaction, results of interview were analyzed. Participants expressed their dissatisfied with the prototype with 24 total icons and 1 icon per page, and the prototype with 48 total icons and 1 icon per page. They generally felt tired to continuously swipe left within the menu, particularly when the number of icon amount was large. Surprisingly, older adults expressed no dissatisfied with the high visual density, although the success rates of taping were reduced. For the reason of older adults' less experience of using smartphones, they preferred the operation of taping rather than swiping, especially on a tiny touchscreen.

The influences of icon amount and visual density on users' satisfaction were tested through MANOVA. There were significant differences in satisfaction among three levels

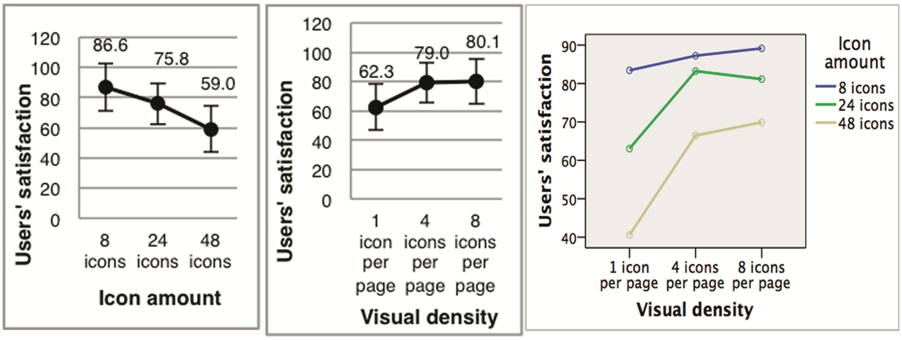

Fig. 6. Users' satisfaction in three levels of icon amount and three levels of visual density

of icon amount ($F(2,126) = 35.2$, $p < 0.001$), and among three levels of visual density ($F(2,126) = 17.9$, $p < 0.001$). Then, multiple comparison was conducted.

As shown in Fig. 6, there were significant differences between icon amount of 8 and 24 icons ($t = 0.001$), 8 and 48 icons ($t < 0.001$), 24 and 48 icons ($t = 0.006$), and between visual density of 1 and 4 icons per page ($t < 0.001$), 1 and 8 icons per page ($t < 0.001$) in users' satisfaction. The results indicated that older adults were sensitive to the increasing numbers of icon amount. Also, they were dissatisfied with the visual density of 1 icon per page. And visual density of 4 icons per page had no significant difference with 8 icons per page in users' satisfaction.

4.3 Interrelations Between User Characteristics, Performance, and Satisfaction

A further consideration refers to the impact of user characteristics, performance and satisfaction (shown in Table 1), which of them might possibly cause or at least interact with each other on tiny touchscreen. This is especially meaningful in the aging group, as aging is going along with the decreases in body flexibility and vision.

Age, vision and finger dexterity did not show meaningful correlations with efficiency and satisfaction. However, there was no surprised that age was significantly correlated to vision and finger dexterity. Further, vision showed significant correlation with effectiveness, meant that reduction of vision led to low success rates of taping and solving tasks.

The success rates of taping were significantly correlated to efficiency (number of swiping left, number of swiping right and tasks completion time). However, the success rates of solving tasks did not affect the efficiency and satisfaction in the different display conditions, which was confirmed by the fact that all tasks were easy to solve. Three aspects of efficiency had significant correlations with each other respectively, showed that with higher efficiency the tasks were solved faster and simpler. Also, satisfaction showed to have a similar effect on efficiency than the success rates of taping, which might mean that older adults judged the prototypes in accordance with their feeling of efficiency.

Table 1. Results of Pearson correlation analysis of characteristics, performance and satisfaction

Variables 1	Variables 2	r	p
Age	Vision	−0.205*	0.017
	Finger dexterity	−0.241**	0.005
Vision	The success rates of solving tasks	0.181*	0.035
	The success rates of taping	0.192*	0.026
The success rates of taping	The number of swiping left	−0.191*	0.027
	The number of swiping right	−0.283**	0.001
	Task completion time	−0.306**	<0.001
The number of swiping left	The number of swiping right	0.597**	<0.001
	Task completion time	0.897**	<0.001
	Users' satisfaction	−0.643**	<0.001
The number of swiping right	Task completion time	0.635**	<0.001
	Users' satisfaction	−0.408**	<0.001
Task completion time	Users' satisfaction	−0.639**	<0.001

Note: *Correlation is significant at the 0.05 level (2-tailed)
** Correlation is significant at the 0.01 level (2-tailed)

4.4 Users' Gestures of Using Smartwatches

The final consideration refers to analysis users' gestures of using the smartwatch. Users' gestures were consisted of the posture of the wrist with a smartwatch, the finger gesture of taping and swiping, and hand gesture of the watch operation.

Regarding the posture of wrist, most of older adults wore the smartwatch on the back of their wrists in the experiment. But three of them wore the smartwatch on the front of the wrist. One participant changed the posture to wear the watch on the front of his wrist in the experimental process. The results of interview showed that they used to wear watches like this and they thought this posture was more suitable for looking and operating the watches. However, correlation analysis indicated that wearing the watch on the front of wrist had low success rates of solving tasks.

Regarding the finger gesture, older adults used their index finger, middle finger and thumb to tap (and swipe) the smartwatch screen in the experiment. Three participants used middle finger, and two of them changed to use index finger in the experimental process. Only one participant used thumb and then changed to use index finger eventually. Correlation analysis indicated that participants who taped with middle finger or thumb had low finger dexterity, and tended to wear the watch on the front of the wrist. However, there were no reasons showed that finger gesture had significant effect on performance and satisfaction.

Regarding the hand gesture, approximately half of participants closed their hands when taping the smartwatch screen with finger, especially middle finger. Positive correlations were found between hand gesture and finger gesture to support that view. There also was significant correlation between hand gesture and posture of wrist, showed that participants who wore the watch on the front of wrist generally tended to close their hands. Moreover, correlation analysis showed that opening hand had low success rates of taping. One possible reason was that opening hand blocked participants' line of sight. Results of Pearson correlation analysis were shown in Table 2.

Table 2. Results of Pearson correlation analysis of Users' gestures and performance

Variables 1	Variables 2	r	p
The posture of wrist	The finger gesture	0.308**	<0.001
	The hand gesture	0.421**	<0.001
	The success rates of solving task	0.257**	0.003
The finger gesture	The hand gesture	0.352**	<0.001
	Finger dexterity	0.469**	<0.001
The hand gesture	The success rates of taping	0.327**	<0.001

Note: ** Correlation is significant at the 0.01 level (2-tailed)

All the tap of prototypes with three levels of visual density was recorded (shown in Fig. 7). Results showed that participants generally tended to tap the bottom right corner of icons. And some of tap crossed the border of icons, especially the icons located at the right side of the screen. As to the tap in prototypes with 1 icon per page, many point of tap were located outside of the icon, which was a demonstration that older adults may carelessly tap the screen when they failed to swipe. What's more, if the visual density was high, participants would probably carelessly tap the icon located at the bottom right corner of the target icon.

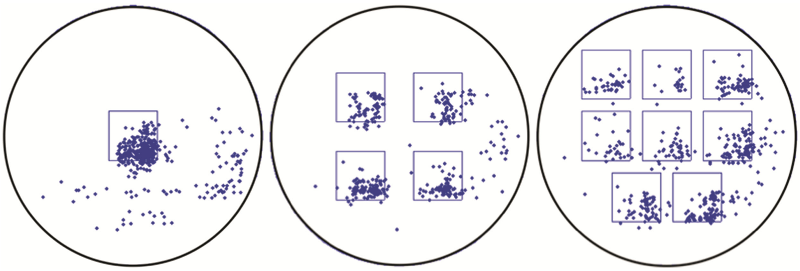

Fig. 7. All point of the tap in prototypes with three levels of visual density (from left to right: 1 icon per page, 4 icons per page, 8 icons per page)

All the swipe of a representative participant was shown in Fig. 8. The difference among the swiping trajectory of three levels of visual density was that participants took obviously fewer number of swiping, no matter swiping left and right, if visual density was higher. And participants only swiped up and down in the prototypes with 1 icon per page. The reason was that they thought low visual density might mean the pages on which they could swipe in both vertical direction and horizontal direction. Moreover, the distance of swiping left (blue line) was longer than the distance of swiping right (red line) in all prototypes. The angle (formed of swiping trajectory and horizontal direction) of swiping right was less than the angle of swiping right.

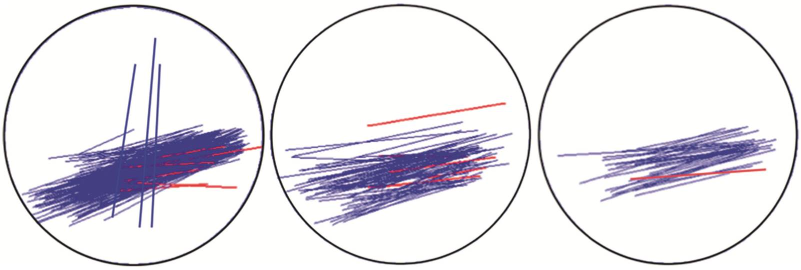

Fig. 8. All swiping trajectory in prototypes with three different levels of visual density (from left to right: 1 icon per page, 4 icons per page, 8 icons per page)

5 Conclusion

Appropriate menu design on the round tiny screen is desired and, hence, this study aims to explore the effect of icon amount and visual density within smartwatches menu on older adults' performance and satisfaction. To achieve this goal, the study proposed different icon amount and visual density of smartwatches and carried out an experiment.

As to the icon amount, menus with 8 icons were better than those with 24 icons and 48 icons in terms of efficiency and satisfaction of older adults. Regarding effectiveness, icon amount of 8 icons was significantly better than 48 icons.

As to the visual density, menus with 1 icon per page were worse than 24 icons per page and 48 icons per page in terms of effectiveness, efficiency and satisfaction of older adults. The menus with 4 icons per page outperformed those with 8 icons per page in effectiveness, but regarding the number of swiping left, 8 icons per pages were better. However, there were no difference between them in other aspects of efficiency and satisfaction.

The results of this study showed which combination of visual density and icon amount was better on the tiny screen (e.g., 40 mm wide and 40 mm high) and which gestures had better performances when using smartwatches. First, a small number of icon amount (e.g., 8 icons) was generally better used in smartwatches menu. In that case, both visual density of 4 icons per page and 8 icons per page had strong qualifying performance. Designers even can used visual density of 1 icon per page if they wanted older adults see icons more clearly. Second, compared with 1 icon and 8 icons per page, the visual density of 4 icons per page might was the best choice if the number of icon amount was large (e.g., 24 icons and 48 icons).

The results indicated that there were relationships between the posture of the wrist with a smartwatch, the finger gesture of taping and swiping, and hand gesture of the watch operation. Wearing the watch on the back of wrist, taping with the index finger and closing hand related to better performance when using smartwatches.

Two limitations of this study also need to be considered. First, icons in the prototypes were different from real icons of smartwatches' applications. Second, the older participants had low education level and cannot be considered as a very representative sample of older users.

Acknowledgment. This work was supported with funding from a National Science Foundation China grant 71401018.

References

1. Chae, M., Kim, J.: Do size and structure matter to mobile users? an empirical study of the effects of screen size, information structure, and task complexity on user activities with standard web phones. Behav. Inf. Technol. **23**, 165–181 (2004)
2. Zhou, J., Rau, P.L.P., Salvendy, G.: Older adults' use of smart phones: an investigation of the factors influencing the acceptance of new functions. Behav. Inf. Technol. **33**, 552–560 (2014)
3. Ziefle, M.: Instruction formats and navigation aids in mobile devices. In: Holzinger, A. (ed.) USAB 2008. LNCS, vol. 5298, pp. 339–358. Springer, Heidelberg (2008)
4. Ziefle, M.: Information presentation in small screen devices: the trade-off between visual density and menu foresight. Appl. Ergon. **41**, 719–730 (2010)
5. Kim, M.J., Yoon, S.K., Choi, J.H.: A study of menu structure on smartwatch for effective navigation - focused on types of menu and color cues. J. Digit. Des. **15**, 395–406 (2015)
6. Perrault, S.T., Lecolinet, E., Eagan, J., Guiard, Y.: Watchit: simple gestures and eyes-free interaction for wristwatches and bracelets. In: the SIGCHI Conference on Human Factors in Computing Systems, pp. 1451–1460. ACM, New York (2013)
7. Knibbe, J., Martinez Plasencia, D., Bainbridge, C., Chan, C.K., Wu, J., Cable, T., Coyle, D.: Extending interaction for smart watches: enabling bimanual around device control. In: The CHI 2014 Extended Abstracts on Human Factors in Computing Systems, pp. 1891–1896. ACM, New York (2014)
8. Cho, H., Kim, M., Seo, K.: A text entry technique for wrist-worn watches with tiny touchscreens. In: The 27th Annual ACM Symposium on User Interface Software and Technology, pp. 79–80. ACM, New York (2014)
9. Komninos, A., Dunlop, M.: Text input on a smart watch. IEEE Pervas. Comput. **13**, 50–58 (2014)
10. Hara, K., Umezawa, T., Osawa, N.: Effect of button size and location when pointing with index finger on smartwatch. In: Kurosu, M. (ed.) Human-Computer Interaction. LNCS, vol. 9170, pp. 165–174. Springer, Heidelberg (2015)
11. Shen, Z., Xue, C., Li, J., Zhou, X.: Effect of icon density and color contrast on users' visual perception in human computer interaction. In: Harris, D. (ed.) EPCE 2015. LNCS, vol. 9174, pp. 66–76. Springer, Heidelberg (2015)
12. Yahoo. How android users interact with their phones. http://yahooadvertisingca.tumblr.com/post/95821374980/yahoo-aviate-has-teamed-up-with-the-data
13. TalkingData. 2014 China mobile Internet report. http://www.talkingdata.com/index/#/datareport/-1/zh_cn
14. Lewis, J.R.: Psychometric evaluation of the PSSUQ using data from five years of usability studies. Int. J. Hum.-Comput. Interact. **14**, 463–488 (2002)
15. Zhou, J., Rau, P.L.P., Salvendy, G.: Use and design of handheld computers for older adults: a review and appraisal. Int. J. Hum.-Comput. Interact. **28**, 799–826 (2012)