Defining and Analyzing a Gesture Set for Interactive TV Remote on Touchscreen Phones

ABSTRACT

In this paper, we recruited 20 participants preforming user-defined gestures on a touchscreen phone for 22 TV remote commands. Totally 440 gestures were recorded, analyzed and paired with think-aloud data for these 22 referents. After analyzing these gestures according to extended taxonomy of surface gestures and agreement measure, we presented a user-defined gesture set for interactive TV remote on touchscreen phones. Despite the insight of mental models and analysis of gesture set, our findings indicate that people prefer using single-handed thumb and also prefer eyes-free gestures that need no attention switch under TV viewing scenario. Multi-display is useful in text entry and menu access tasks. Our results will contribute to better gesture design in the field of interaction between TVs and touchable mobile phones.

Author Keywords

User-elicitation methodology, remote control, user-defined gesture, eyes-free, touchscreen.

ACM Classification Keywords

H5.2. Information interfaces and presentation: User Interfaces – *Interaction styles, user-centered design*.

General Terms

Design, Experimentation, Human Factor.

INTRODUCTION

TV, the media center of smart space at home, had become a terminal for interactive applications such as online shopping and media display. With the development of interactive TVs, traditional push-button based remote controls were not flexible enough to meet the complicated demands of these applications. New remote control technologies based on touchscreen [3, 4, 5] have been designed and evaluated by many studies. However, these researches focused on the efficiency of input (e.g. pointing).

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

UbiComp '13, Sep 8–Sep 12, 2013, Zurich, Switzerland.

Copyright 2013 ACM 978-1-4503-1770-2/13/09...$10.00.

As smarts phones gradually turned into the most pervasive computing devices interacting with smart environments [2], many applications use smart phones as the control device for interactive TV. Apple Remote [1] presents a 7 keys design supporting tap and flick on touchscreens and support copy screen mode. Samsung Remote [9] and UnityRemote [10] are designed as soft-key-based TV remote supporting relative position control of touchscreen. But there is nearly no study about the user-centered gesture design on touchscreens of smart phone in the TV viewing scenario.

In this study, we adopted user-elicitation methodologies [6, 7, 8, 11] to define and analyze a complete user-defined gesture set for interactive TVs based on touchscreens. Firstly we summarized three basic interaction tasks on interactive TV: navigation, text entry and shortcuts obtaining 22 referents. Then following the user-elicitation methodology, 440 user-defined gestures on touchscreen phones from 20 participants were recorded and classified according to the extended taxonomy of 5 dimensions: *form, nature, binding, flow* and *attention*. After analyzed using agreement measure [11], 34 gestures were mapped to 22 referents with minimum consensus-threshold of 4. Finally, despite of the insight of mental model, we presented some findings of the gesture vocabulary and eyes-free gesture design in discussion section. Our results contribute to a user-defined gesture set and insight into users' mental models and also eyes-free gesture design for remote control on touchscreen phones. This work will help designers with their future design in the field of TV and phone interaction.

DEVELOPING A USER-DEFINED GESTURE SET

Understanding the Tasks on Interactive TV

After exploring the functions of interactive TV including playing movie and music, searching the Internet and so on, we obtained the common commands as Table 1 shows. There were three tasks including *navigation*, *text entry* and *shortcut* with totally 22 referents. Other commands such as delete, copy and paste were not considered because of the low usage frequency. But they can be navigated after "Show menu" command performs. We evaluated the coverage of these referents with the UI elements and commands of Samsung smart TV, Apple TV, Google TV and XBMC. It proved that this referent set include all the referents in TV viewing scene.

After obtaining the basic tasks and referents, we designed a lab-based user study following the user-elicitation methodology mentioned by Wobbrock *et al.* [11].

Participants

20 paid participants (age ranges from 20 to 40, M=26.7; 6 females) with different educational background volunteered for user-elicitation design (12 students from different disciplines: computer science (6), art (3), finance (2) and biology (1); and 8 staffs from different industries: IT (3), education (2), industrial design (2) and music (1)). All of them had the experience of touchable mobile devices.

Procedure

The study was conducted in a lab that was decorated as a living room, with a 46 inches TV set at 1920×1080 resolution and a sofa 8.2 feet (2.5 meters) in front of the TV. The participants seated on the sofa facing the TV screen with an iPhone 4S in hand. Firstly we introduced and explained them the referents presenting a general view to avoid the design misunderstanding. Then, the participant was asked to carry out an appropriate gesture on phone screen for each command (randomly arranged), with the whole process thinking aloud. . When they finished design a gestures, we triggered the feedback of the corresponding referent. The design procedure was recorded with a video camera. The video data was later manually annotated to calculate thinking and articulation time of each gesture.

After the design of each referent, the participant was asked to give two 7-point Likert scales (-3=*strongly disagree*, 0=*undecided*, 3=*strongly agree*) concerning the suitability of the gesture (*The gesture is a good match for the intended purpose*) and ease of perform of the gesture (*The gesture is easy to perform*). With $22 \times 20 = 440$ recorded gestures (see Table 1), we obtained 107 different kinds of gestures.

Classification of Gestures

User-elicitation methodologies have been conducted to design several solutions including multi-touch gestures [11], mobile motion gestures [6, 8] and other multimodal gestures [7] for specific scenarios or devices. However, compared with these related works, our study focused on multi-screen scenario with a small touch and display screen in hand and a TV screen away from users. In this scenario attention switch between displays has influence on the user experience [12] and is an essential dimension classifying gestures or operations. So we extended the taxonomy into five-dimensions: *form, nature, binding, flow* and *attention*.

In the *form* dimension, the gestures are distinguished between 'static' and 'dynamic' 'pose' and 'path' of one hand. 46.6% (205 of 440) of the gestures are one-finger gestures (touch and path) with single-handed thumb (154) or two-handed index finger (51) contacting with touchscreen. Two-finger gestures (124 of 440 = 28.2%) are preferred when the user cannot come up with a reasonable one-finger gesture such as 'fast forwards' (swipe to the right with index and middle fingers together). multi-finger gestures (50 of 440 = 11.4%) were designed for some referents with high complexities like 'Task switch' (three finger together swipe to the bottom). The other 13.9% (61 of 440) of the gestures are classified as 'static (dynamic) pose (and path)' such as five fingers' splay on touchscreen representing 'Task switch' command.

Referents	Complexity Mean	Complexity SD	Referents	Complexity Mean	Complexity SD
Navigation			Up	1.33	0.58
Select	1.33	0.58	Down	1.33	0.58
Pointing	1.67	1.15	Show menu	1.33	0.58
Scroll	3.00	1.00	Zoom in	2.33	0.58
Menu Access	4.00	1.00	Zoom out	2.33	0.58
Text Entry			Home	2.67	1.15
Number	3.67	0.58	Back	3.33	1.00
Character	4.67	0.58	Forward	3.33	1.00
Shortcut			Fast forwards	3.33	1.52
Play	1.00	0.00	Fast backwards	3.33	1.52
Pause	1.00	0.00	Help	4.67	0.58
Next	1.33	0.58	Task switch	5.00	0.00
Previous	1.33	0.58	**Mean**	**2.61**	**0.71**

Table 1. 22 commands are listed in a predefined order grouped by tasks. Each command's conceptual complexity was rated by the authors (1=*simple*, 5=*complex*).

In the *nature* dimension, gestures were separated into four sets: symbolic (e.g. writing a question mark on the screen), physical (e.g. swipe down), metaphorical (e.g. clicking an imaginary button), or abstract (in case of an arbitrary mapping). Similar results were observed as Wobbrock *et al.* and Kühnel *et al.* mentioned. The *nature* of the gestures is influenced by the conceptual complexity of the referents ($F_{3,440} = 43.3, p < .001$). Besides, the nature of the gestures has significant effect on gesture design ($F_{3,440} = 70.6, p < .001$) time and number of matches ($F_{3,134} = 12.8, p < .001$). We found that physical gestures led the shortest design time (M = 4.34s, SD = 3.65s) and abstract gestures the longest design time (M = 14.41s, SD = 10.62s).

In the *binding* dimension, gestures were classified into 5 sets: object-centric (considering object location features), time-dependent (considering the time (speed) features without object location features), world-dependent (considering the location feature without object location features), world-time-independent (can occur anywhere) and mixed dependencies (combination of several *binding* sets). According to the participants' thinking aloud design, we found the reason why *world-time-independent* gestures are preferred (54.8%) is the 'eyes-free' factor [12]. 87.5% of the *world-dependent* gestures were designed because of the soft-key-based design on phone screens.

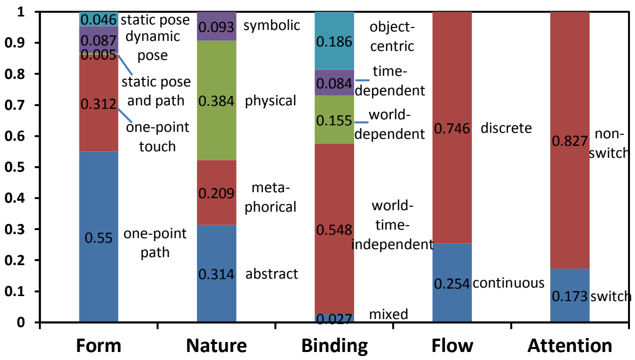

Figure 1. Percentage of gestures designed in each category.

Flow dimension distinguishes the gesture between continuous (response occurs while the user acts) with a small percentage (25.4%) and discrete (response occurs after the user acts). *Attention* dimension distinguishes whether the tasks or commands need attention switch (17.3%) between two screens or focus on one display ignoring the other.

RESULT AND ANALYSIS OF GESTURE SET

We adopted the agreement measure A_r and A method of Wobbrock *et al.* [11]:

$$A_r = \sum_{P_i \subseteq P_r} \left(\frac{|P_i|}{|P_r|}\right)^2, A = \frac{\sum_{r \in R} A_r}{|R|} \quad (1)$$

Eq.1 computes the agreement A_r of each referent r of the 22 commands and A of whole referents R of the gesture set. P_r is the set of proposed gestures for referent r and P_i is a subset of identical gestures from P_r. The agreement for this study is as Figure 2 shows. The complete gesture set is shown in Figure 3.

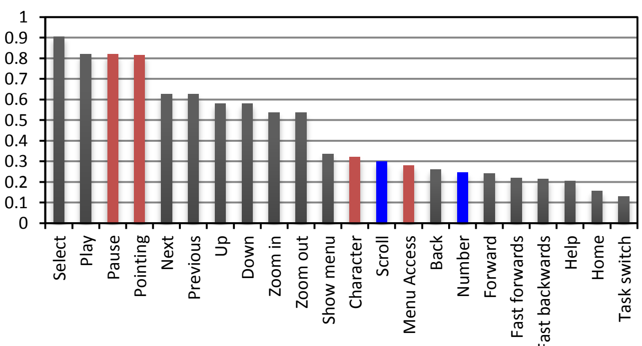

Figure 2. Agreement A for each referent in descending order. The pink color represent navigation task, blue color of text entry and gray color of shortcut.

Similar findings of Wobbrock *et al.* [11] were obtained when analyzing planning time, articulation time and subjective performance.

- The conceptual complexity has a significant correlation with average gesture planning time ($r = .64, F_{10,440} = 15.4, p < .001$). But there is no significant effect of complexity on articulation time ($F_{10,440} = 0.43, p = ns$). In general, the more complex the referent is, the longer the more time participants spend for designing the gesture. It took the participants 18.50s (SD=15.84s) to design a gesture for the most complex referent (Task switch) while simple referents (play/pause) took participants 1.76s (SD=2.65s) on average to design a gesture.
- The referents' conceptual complexities correlated strongly with the rating on suitability of gestures ($r = -.77, F_{10,440} = 45.2, p < .001$) and ease to perform of gesture ($r = -.57, F_{10,440} = 10.2, p < .001$). The suitability rating ranges from ($M = 0.14, SD = 1.15$) of 'Home' referent to ($M = 2.21, SD = 0.65$) of 'Play/Pause' referents. And the ease to perform rating ranges from ($M = -0.34, SD = 1.65$) of 'Task Switch' referent to ($M = 1.64, SD = 1.27$) of 'Play/Pause' referents.

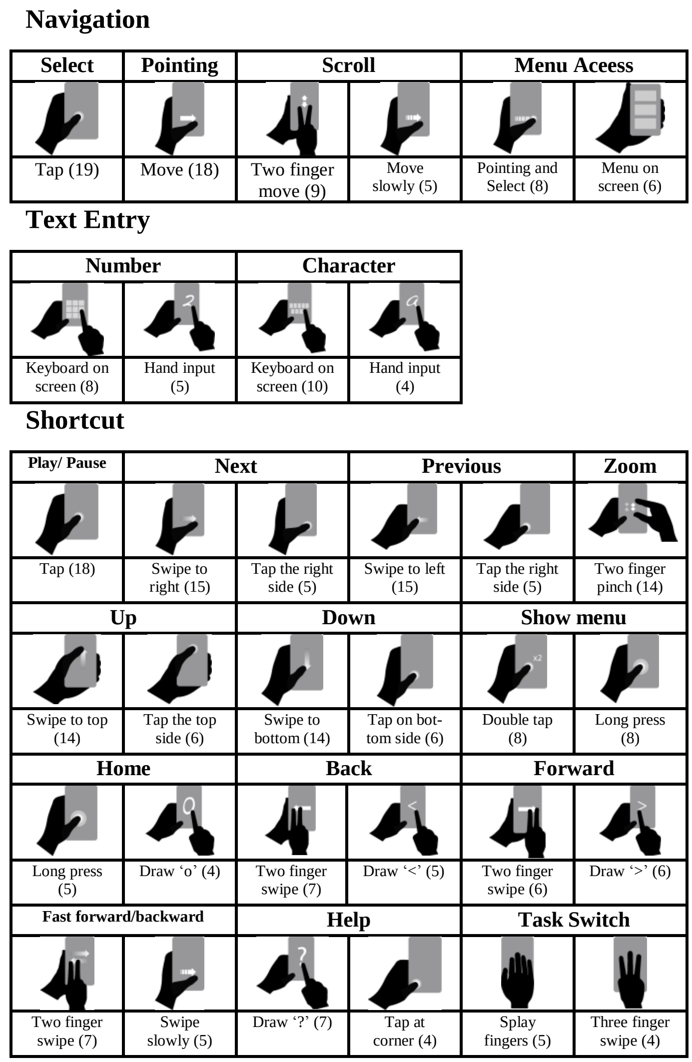

Figure 3. Referent-gesture mapping of different tasks with a consensus-threshold of four. The number represents the number of matches. Designers can give different gesture designs following this figure and are expected avoiding the conflicts between gestures.

- The planning time has a significant effect on suitability rating ($F_{176,440} = 8.2, p < .001$) and ease to perform rating ($F_{123,440} = 3.2, p = .002$).

DISSCUSSION

In this section, we discuss the gestures set analysis, insight of mental models and preference of eyes-free gestures.

Gesture Set Analysis

Unique and overlap of gesture design. We assigned 1 to 2 gestures for each referent with a consensus-threshold of four. The gestures are not unique across referents (Tap for both "Select", "Play" and "Pause"). However, some overlaps may occur in specific scene (Two finger swipe for both "Scroll", "Back" and "Forward" of watching movie).

Switch between gestures. Some gestures need switch to be activated. "Stroke-based number input" need to be activated to input number. Otherwise it would be conflict with one-point path gestures (number 1 and swipe to bottom).

Mental Model

Design reversible gestures with dichotomous referents. We get the same result as Wobbrock *et al.* mentioned that people generally design reversible gestures for dichotomous referents (next/previous, up/down, zoom in/zoom out).

Prefer single-handed thumb interacting with touchscreen. Participants give preference to single-handed thumb contacting with touchscreen phones (154 of 440 = 35.0%). Only one participant (female) used index finger of another hand designing one-point path and touch during the whole designing procedure. Two participants (1 female and 1 male) used both single-handed thumb and two-handed index finger designing one-point path and touch. And one participant suggested distinguishes among fingers.

Low conceptual complexity leads to common gesture and high complexity to mass gesture. People tend to design a unique gesture for referent with high conceptual complexity. Only 2 gestures was designed for "Play", "Pause", "Next", "Previous", "Up", "Down", "Zoom in" and "Zoom out" that have a low conceptual complexities. But 11 gestures were designed for "Task switch", 9 for "Help" and 8 for "Fast backward" and "Home" with higher conceptual complexities.

Multi-screen in TV Viewing Scene

Eyes-free gestures are preferred. We extended the gesture classification from 4 to 5 dimensions adding *attention* to test the eyes-free factor. 82.7% of the gestures participants designed are eyes-free gestures that need no attention switch between the TV screen and touchscreen. This finding confirm the findings of Yi *et al.* [12] that people adopt eyes-free interaction to pay more attention to the surrounding activities.

Display on touchscreen is preferred in menu access and text entry tasks. Near half of the participants prefer soft keyboard or corresponding menu on touchscreen that can be "reached" directly according what P2 and P8 said. One participant presented a soft-key-based design laying all the commands on the phone screen. But he said, "Most frequently used keys should be placed on the first page and the key used frequently should have a larger size."

CONCLUSION AND FUTURE WORK

In this study, we presented and analyzed the user-defined gesture vocabulary of interactive TV remote on touchscreen phones based on participants' agreement of 440 gestures. The analysis of gesture set, insight into the mental models and rationality of eyes-free gestures was presented under multi-screen TV viewing scene. This work was expected to contribute to the future design on touchscreen-based remote control. More and more sensors had been embedded into the mobile phone making motion gestures easy to perform. And our current study only presented the user-designed gesture set on touchscreen of mobile phones without usability test. In the future, combination of the touch and motion input modalities to present highly usable gestures and task-based usability test are promised to be conducted.

REFERENCES

1. Apple Remote – The all-new Remote.
   http://www.apple.com/apps/remote/
2. Ballagas, R., Rohs, M., and Sheridan, J. G. Mobile phones as pointing devices. In PERMID 2005.
3. Choi, S., Han, J., Lee, G., Lee, N., and Lee, W., RemoteTouch: Touch-Screen-like interaction in the TV viewing environment. In *Proc. CHI'11*, 393-402.
4. Cox, D., Wolford, J., Jensen, C. and Beardsley, D. An evaluation of game controllers and tablets as controllers for interactive TV applications. In *Proc. ICMI'12*, ACM Press (2012), 181-188.
5. Enns, N. R., and MacKenzie, I. S. Touchpad-based remote control devices. In *Proc. CHI'98*, ACM Press (1998), 229-230.
6. Kühnel, C., Westermann, T., Hemmert, F., Kratz, S., Müller, A., and Möller, S. I'm home: Defining and evaluating a gesture set for smart-home control. *IJHCS*, *69*(11), 693-704.
7. Morris, M. R. Web on the wall: insights from a multimodal interaction elicitation study. In *Proc. ITS'12*, ACM Press (2012), 95-104.
8. Ruiz, J., Li, Y., and Lank, E. User-defined motion gestures for mobile interaction. In *Proc. CHI'*11, ACM Press (2011), 197-206.
9. Samsung Remote – Remote App of Samsung TV.
   https://play.google.com/store/apps/details?id=com.samsung.remoteTV
10. UnityRemote – Universal Remote for iPad and iPhone
    http://www.gear4.com/product/_/29/unityremote.
11. Wobbrock, J. O., Morris, M. R., and Wilson, A. D. User-defined gestures for surface computing. In *Proc. CHI'09*, ACM Press (2009), 1083-1092.
12. Yi, B., Cao, X., Fjeld, M., and Zhao, S., Exploring User Motivations for Eyes-free Interaction on Mobile Devices. In *Proc.CHI'12*, ACM Press (2012), 2789-2792.