國立臺灣大學電機資訊學院資訊網路與多媒體研究所
碩士論文
Graduate Institute of Networking and Multimedia
College of Electrical Engineering & Computer Science
National Taiwan University
Master Thesis

智慧型眼鏡在公眾場合中的使用者定義遊戲操作
User-Defined Game Input for Smart Glasses in Public Space

王瀚宇
Han-Yu Wang

指導教授：陳彥仰博士
Advisor: Mike Y. Chen, Ph.D.

中華民國 104 年 6 月
June, 2015

國立臺灣大學碩士學位論文
口試委員會審定書

智慧型眼鏡在公眾場合中的使用者定義遊戲操作
User-Defined Game Input
for Smart Glasses in Public Space

本論文係王瀚宇君 (R02944002) 在國立臺灣大學資訊網路與多媒體研究所完成之碩士學位論文，於民國 104 年 6 月 2 日承下列考試委員審查通過及口試及格，特此證明

口試委員：

所　長：

誌謝

首先誠摯的感謝指導教授陳彥仰博士，在研究所期間的悉心指導使我得以一窺人機互動領域的門徑。討論時循循善誘的引導方式，讓我在研究上獲益匪淺。在實驗室的這段時間老師以身作則的示範了許多世界級的做事方式是我將來學習的典範。

本論文的完成另外亦得感謝童盈超、許鈞彥、邱詩雅、林哲緯、吳沛容及林欣穎等人的大力協助，因為有你們的體諒與幫忙，才使得本論文能夠完整而嚴謹。

在台大人機互動實驗室的兩年，我有幸跟世界上最頂尖的一群天才創造了許多回憶，認真的學術討論、言不及意的鬼扯淡、深夜的大腸花、趕死線時的革命情感、垃圾清不掉的叫罵聲.... 這些是我們之間不可取代的羈絆。感謝眾位學長姐、同學、學弟妹的共同砥礪 (墮落？)，你/妳們的陪伴讓我兩年的研究生活變得多采多姿。

感謝許富傑、蔡銘綸、梁翔勝、蔡明璋、李孟翰、王正堯、王偉翰、陳厚任、鄭達陽、詹力韋、黃大源、梁容豪學長及簡妙恩、黃易學姊，你們的悉心指導與金玉良言，對我這兩年的研究生活影響深遠。也感谢李柏緯、何駿銘、張巧慧同學的幫忙，恭喜我們順利度過這兩年。實驗室的王權、周冠廷、修敏傑學弟、林敬沂學妹們當然也不能忘記，你/妳們聽我講垃圾話辛苦了。

女朋友王玉玲在背後默默支持更是我前進的動力，相信沒有妳的陪伴這兩年的生活會是很不一樣的光景。

最後再次感謝我在研究所的這段時間所遇到的所有人，是你們成就了這篇論文的每字每句。

摘要

智慧型眼鏡（如：Google Glass）與傳統的主機或行動遊戲平台不同，具備有隨時都可遊玩的特性，有創造無處不在的遊戲體驗的潛質。然而現行的智慧型眼鏡遊戲操作方式侷限在現有的軟硬體感應科技上。為了瞭解使用者真正想要的設計方向，在本篇論文中我們跳脫現行科技限制，探索了使用者在公眾場合中喜歡的智慧型眼鏡遊戲操作方式。

我們找了二十四名受測者執行了一場使用者定義遊戲操作實驗（User-Defined Game Input Study），實驗範圍囊括了十七種常見的遊戲操作、三種類別的人機互動方式跟兩種不同型式的智慧型眼鏡，最後共執行了兩千四百四十八次的遊戲操作嘗試。

根據我們的結果表示，相較於手持操作器，使用者有顯著性差異的較喜歡使用非觸碰式的操作（如：空中手勢）。在非手持的觸碰式操作中，使用者最喜歡的互動操作位置是手掌而不是穿戴式裝置 (51% vs 20%)。除此之外還發現在公眾場合中使用者會考慮社會認可的問題 (Issue of Soical Acceptance) ，所以較喜歡使用不引人注意的操作方式，導致在使用空中手勢時使用者喜歡的操作範圍是在軀體 (Torso) 前方而非面前 (63% vs 37%)。

Abstract

Smart glasses, such as Google Glass, provide always-available displays not offered by console and mobile gaming devices, and could potentially offer a pervasive gaming experience. However, research on input for games on smart glasses has been constrained by the available sensors to date. To help inform design directions, this paper explores user-defined game input for smart glasses beyond the capabilities of current sensors, and focuses on the interaction in public settings. We conducted a user-defined input study with 24 participants, each performing 17 common game control tasks using 3 classes of interaction and 2 form factors of smart glasses, for a total of 2448 trials. Results show that users significantly preferred non-touch and non-handheld interaction over using handheld input devices, such as in-air gestures. Also, for touch input without handheld devices, users preferred interacting with their palms over wearable devices (51% vs 20%). In addition, users preferred interactions that are less noticeable due to concerns with social acceptance, and preferred in-air gestures in front of the torso rather than in front of the face (63% vs 37%).

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Smart glasses provide always-available displays and offer the opportunity for instantly available information and pervasive gaming experiences. Compared to game consoles and mobile gaming devices, smart glasses do not have touchscreens and currently do not support handheld controllers specifically designed for gaming. Current smart glasses, such

Figure 1.1: A study participant performing an in-air gesture to drag an object seen through the immersive smart glasses in a public coffee shop.

as Google Glass and the Epson Moverio, support input via voice, touchpads, cameras, gyroscopes, accelerometers, and GPS. Games designed specifically for Google Glass [27] utilize these sensors as game control. For example, “Clay Shooter” utilizes the user’s voice to trigger a shotgun, and “Shape Splitter” detects in-air gestures via the built-in cameras. For the Epson Moverio glasses, wired trackpads are used as handheld inputs.

To better inform the interaction design of games for smart glasses, we aimed to explore the design space without being constrained by the capabilities of current sensors. We used the guessability study methodology [42], and presented the *effects* of game controls to the participants in a real-world, public environment. We then elicited what the participants felt was the most appropriate *causes* to invoke the corresponding effects.

For input tasks, we analyzed 90 popular games to identify the game controls used by more than one game, which resulted in a set of 17 tasks. We also explored the form factors of smart glasses displays, and included both types in the study: 1) *immersive*, with display content spanning the user’s field of view (e.g. Epson and Sony’s smart glasses), and 2) *off-to-the-side*, with display content in the corners of the user’s field of view (e.g. Google Glass).

In order to compare different types of interaction while keeping the experiment tractable, we grouped the different types of input into the following 3 classes:

- *handheld*: input types that make use of handheld controllers, such as smartphones and the wired trackpads used by Sony’s SmartEyeglass and Epson’s Moverio glasses.

- *touch*: non-handheld touch input, such as gesturing and tapping on body surfaces, and touch-sensing wearable devices (e.g. smart rings, watches, and glasses). These provide tactile feedback.

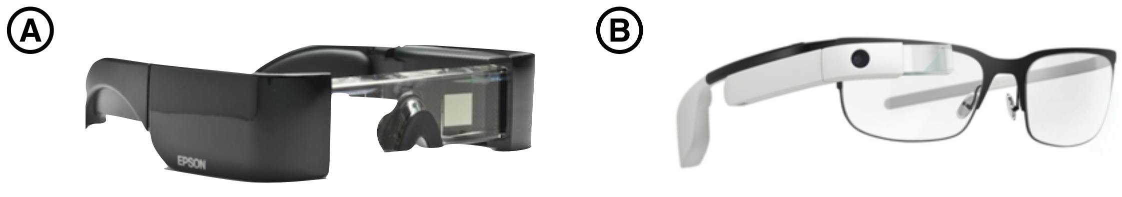

Figure 1.2: (A) Epson Moverio, (B) Google Glass.

- *non-touch*: non-handheld, non-touch input, such as in-air gestures, head/body movement, and voice recognition. These do not have tactile feedback.

We recruited 24 participants and asked them to wear the two form factors of smart glasses in a coffee shop. On-screen instructions prompted participants to perform each of the 17 common game control tasks using the 3 classes of input types. For each game control task, form factor, and input type, participants first explored all possible interactions they could think of, then reported the one they most preferred. After completing the 3 types of interactions for that task and form factor, they then rated their preferences for 3 interactions. Overall, each participant reported 102 interactions, for a total of 2448.

We collected quantitative and qualitative data through video analysis, preference ratings, and interviews. Our key observations are as follows:

- Participants significantly preferred non-handheld, non-touch interactions over handheld interactions (3.81 vs 3.68 on a 5-point Likert scale, $p<0.01$).
- For touch input without using handheld devices, users preferred interacting with their body surface over wearable devices (80% vs 20%), and the most frequently used body surface was the palm (51%).
- Participants preferred interactions that are more subtle due to concerns with social acceptance. Also, participants preferred using in-air gestures in front of the torso than in front of the face (63% vs 37%), even though those gestures were reported to be less intuitive and less precise.
- There is a significant mismatch between participants' preferred input methods and those supported by the current smart glasses. For example, less than 2% of the participants used voice and less than 2% of the participants used touch input on the smart glasses – which are Google Glass' two primary input methods. In addition, current cameras can only detect in-air gestures in front of users' faces, missing most 63% of the gestures performed.

The contribution of this paper are as follows: (1) the first quantitative and qualitative characterization of user-defined input for games on smart glasses, including a taxonomy,

(2) set of user-defined input for common game tasks, which is reflective of user behavior.
(3) insight into users' mental models when playing smart glasses games in a public space, and an understanding of implications for mobile input technology and user interface design. Our results will help designers create better smart glasses experience informed by user behavior.

Chapter 2

Related Work

2.1 Game Input

There are many previous works exploring new kinds of game controls and pushing the limit of game design. Vickers et al.[40] showed the possibility to use eye gestures as game inputs; Christian et al.[8] provided novel techniques for users to interact with games by head-gesture; Harada et al.[18] and Sporka et al.[36] both indicated that the voice input greatly expanded the scope of games that could be played hands-free and just counted on voice input; Baba et al.[4] presented a game prototype which treated skin contact as controller input; Nacke et al.[30] even considered using biofeedback (including EMG, EDA, EKG, RESP, TEMP) as game input methods; Hsu et al.[21] compared different game inputs, including head gestures, voice control, handheld controller, joysticks, eye winking and glass touchpad, for First-Person Shooter(FPS) games on smart glasses.

2.2 Mobile Input Technology

Some works related to mobile systems had defined designer-made input methods. These systems could be divided into two main categories, *touch* and *non-touch* inputs.

Harrison et al.[19] created *OmniTouch*, a wearable depth-sensing and projected system that enables interactive multitouch applications on any surface of the user's body. Moreover, *Skinput*[20], a technology that appropriates the human body for acoustic transmission, and allows the skin to be used as an input surface. Chan et al. presented FingerPad[7], a nail-mounted device that turns the tip of the index finger into a touchpad, allowing private and subtle interaction while on the move. Baudisch et al.[17] illustrated a concept of imaginary interface with sensing several gestures on the user's palms. Recently, Serrano et al.[35] explored the use of *Hand-to-Face* input to interact with head-worn displays(HWD) and provided a set of guidelines for developing effective Hand-to-Face interactions based on two main factors they found, social acceptability and cultural effect.

Kim et al.[25] developed a wrist-worn architecture, which supports discrete gesture recognition with reconstructing a 3D hand model in the air. Similarly, Jing et al.[22] implemented *Magic Ring*, a finger ring shaped input device using inertial sensors to detect subtle finger gestures; Colaço et al.[9] built a head-mounted display, *Mime*, sensing 3D gestures in front of the user's eyes.

2.3 Gestures in HCI

Gesture-based interfaces are already common in a variety of application domains such as gaming, virtual or augmented reality and mobile devices[24]. Aigner et al.[1] conducted a study of human preferences in usage of gesture types for HCI and indicated that, depending on the meaning of the gesture, there is preference in the usage of gesture types; Nielsen et al.[31] pointed out some important issues in choosing the set of gestures for the interface from a user-centred view such as the learning rate, ergonomics, and intuition; Grijincu et al.[16] presented a video-based gesture dataset and a methodology for annotating video-based gesture datasets; Recently, Piumsomboon et al.[32] have developed a user-defined gesture set for augmented reality applications. In our work, we focus on exploring relevant input gestures for gaming on smart glasses in public space.

2.4 User Elicitation Studies

User-elicitation studies are a specific type of participatory design methodology that involves end-users in the design of control-sets[29]. These studies had been used to design user interfaces of various types including multi-touch gestures on small and large surfaces[2, 43] and multi-modal interactions [29, 26]. There is also some evidence that user-defined control sets are more complete than those sets defined solely by experts[33, 43].

In a user-elicitation study, users were shown referents (an action's effects) and were asked to demonstrate the interactions that resulted in a given referent[43]. In this work, we draw upon the user-elicitation methodology to identify user expectations and suggestions for smart glass gaming.

Chapter 3

Developing a User-Defined Game Input Set

3.1 Overview

We developed a user-defined game input set by having 24 participants perform game tasks with smart glasses. To avoid bias from visual hints[11], no elements specific to PCs, consoles and mobile games were shown. Similarly, no specific game title was assumed. Instead, participants acted in a simple blocks world of geometry shapes or in the shape of a basic human avatar. Each participant saw the effect of a game input (e.g. an object moving left and right) and was asked to perform the game input action he or she thought would use to cause that effect (e.g. performing an in-air gesture to drag the object left and right, see Figure 1.1).

Seventeen game tasks were presented, and game inputs were elicited for three different interaction methods (*handheld*, *touch*, *non-touch*) with 2 smart glasses (Google Glass, Epson Moverio). The system did not attempt to sense the user's input action, but we used a camera to record the whole process. Participants used the think-aloud protocol and were interviewed about the input details. They also provided subjective preference ratings.

The final user-defined game input set was developed in light of the *agreement* found in the participants' preferred input action for each game task. The more participants that used the same action for a given task, the more likely that input action would be assigned to the task. In the end, our user-defined game input set emerged as a surprisingly consistent collection founded on actual user behavior.

3.2 Interaction Methods

In our study, we asked users to define three input manners to satisfy three interaction requirements individually in each task. These three interaction types, classified according to familiar interactions explored by previous works, were *handheld*, *touch*, and *non-touch*. *handheld*, one of these types, required users to create a game input by interacting with common portable handheld devices, mobile phones. Another method was *touch*, which asked users to design an input action by touching any skin, clothes or accessories on their own bodies. The last method, *non-touch*, was that users were asked to define an input method without touching any tangible object, such as, moving eyeballs, rotating their heads, voice control or in-air gestures.

3.3 Game Tasks

Casual game is one of the game categories with the most players[13], and it is shown high potential in public gaming[34, 5]. We chose top 90 casual games[38] from existing platforms, including PCs, consoles and mobile games (30 games for each) by crawling and analyzing the sale and download count data from famous gaming websites[3, 39, 37, 14]. We invited 3 experienced game developers to review these top 90 casual games. In these games, they found 26 game tasks in total, and removed 9 tasks which were only used once in specific games. Finally, we got a set of general casual game task (shown in Table 3.1) with 17 tasks, which can completely support 90% of our top casual games.

#	Task	Used in Famous Game
1	Select single from many	Clash of Clans, Plague Inc.
2	Vertical menu	Puzzle&Dragon, PeggleHD
3	Horizontal menu	Clash of Clans, PeggleHD
4	Move left and right	Temple Run, Super Mario
5	Move in 4 directions	1943, RaidenX
6	Switch 2 objects	Candy Crush, Bejeweled
7	Move object to position	World of Goo, The Sim
8	Draw a path	Draw Something, P&D
9	Throw an object (in-2D)	Angry Birds, PeggleHD
10	Follow the beats	RockSmith, Guitar Hero
11	Rotate an object (Z-axis)	Zuma, PeggleHD
12	Rotate an object (Y-axis)	Spore, The Sim
13	Avatar jump	Temple Run, Super Mario
14	Avatar 3D move	Spore, Tintin
15	Avatar attack	Minecraft, Terraria
16	Avatar squat	Temple Run, Minecraft
17	Control 3D viewport	The Sim, Spore

Table 3.1: Summary of our general casual game task set. We named several famous games which use these tasks.

3.4 Form Factor of Glasses

We explored the form factors of smart glasses displays, and included both types in the study: 1) *immersive*, with display content spanning users' field of view (e.g. Epson Moverio), and 2) *off-to-the-side*, with display content in the corners of users' field of view (e.g. Google Glass). The display of the Epson Moverio is located in front of the user's eyes with 960×540 resolution[12]. And Google Glass locates its display above the user's right eye with 640×360 resolution[15] (see Figure 1.2).

3.5 Participants

We recruited twenty-four participants with an equal male-female ratio for our study. Their average age was 23.2 (*sd* = 2.72). All participants are right-handed and none of them had past experience with smart glasses usage. About their gaming experience, according to our investigation, 14 users were daily game players, 9 were weekly players and 1 was a monthly player. Participants spent 1.36 hours (*sd* = 0.89) on average to play games one time. Moreover, 58% of them indicated that their main gaming platforms were mobile phones, 38% were on PCs, and only 4% were on consoles. Another important factor of the gaming experience is the user's familiarity with game controllers. The results showed that average familiarity scores were 2.50 for gamepad (*sd* = 1.24), 3.96 for touchscreen (*sd* = 0.62) and 4.22 for keyboard and mouse (*sd* = 0.74) on a 5-point Likert Scale for degree of familiarity (1 means very unfamiliar, 5 means very familiar).

3.6 Environment

According to the previous works[41, 28], the social acceptability of mobile-input was influenced by whether participants believed a bystander could interpret the intention of the input action. Therefore, to provide a game input set suited for a real-world environment, we chose a Starbucks cafe near our college. The visitor flow of the cafe, on average, was 72.5 persons per hour. In our investigation, participants indicated that the cafe was comparatively a public space with average 4.17 points (*sd*=0.65) on a 5-point Likert Scale for degree of field publicity (1 means very private, 5 means very public).

3.7 Procedure

Participants wore two different glasses (Google Glass and Epson Moverio) and our software randomly presented 17 game tasks (Table 3.1) to participants. For each game task, participants performed an input action in 3 different interaction methods (*handheld*, *touch* and *non-touch* interaction). The study was conducted using a counterbalanced measures design, alternating the glass's form and the interaction method. After each game input, participants were shown a 5-point Likert scale concerning subjective preference and conducted a short interview about input detail. With 24 participants, 17 game tasks, 2 glass forms and 3 interaction methods, a total of $24 \times 17 \times 2 \times 3$ = 2448 game input actions were made.

Chapter 4

Results

Our results include game input taxonomies, a user-defined game input set, user ratings, subjective responses, and qualitative observations for each interaction method.

4.1 Preference Between Interaction Methods

Table 4.1 shows the average rating of 3 interaction methods. Three interaction types had a significant rating difference ($F_{0.05}(2, 2445)=4.61$, p = .01). We found that the user rating preference for *non-touch* was significantly higher than for *handheld* (p = .009). And we didn't find a significant difference between *touch* and *non-touch* (p = .688).

In the interview conducted after the rating, we asked why users gave *handheld* a lower preference score. The general reason was that users had to take their controller, e.g. phone, out of their pocket first. Users thought the handheld controller was not always-available and was not hands-free compared to the other interaction methods in this study. After analysing the video recording, we found that most participants simply used the handheld

Method	Mean	Std.	L.Bound	U.Bound
handheld	3.68	0.79	3.63	3.74
touch	3.77	0.81	3.72	3.83
non-touch	3.81	0.90	3.75	3.87

Table 4.1: Summary of user preference of 3 different interaction methods, it provides mean value, standard deviation, and 95% confidence interval for mean (Lower Bound and Upper Bound).

device as a trackpad. In addition, the user behavior and mental model were similar to the previous work by Liang et al.[26]. Considering the users' preferences and the reasons mentioned above, our report will focus on the results of *touch* and *non-touch* interaction.

4.2 Behavior with Different Form Factor of Glasses

In our study, for each of the two glasses, there are 1224 game input pairs with the user, task and interaction method. We found 119 pairs of game input (9.72% of all) were designed differently with distinct smart glasses forms. The influence of game input in each interaction method was 1.22% for *handheld*, 7.35% for *touch* and 20.59% for *non-touch*.

While using *non-touch* as interaction method, users who designed distinctive game input action mentioned that they were eager to use direct control and perfom a in-air gesture in front of the screen with the Epson Moverio. However, it is difficult to perform the same input with Google Glass because of the small screen size. On the other hand, users' reasons to define different input with different glasses for *touch* and *handheld* interaction methods seem to be random as users were unable to motivate their choice to define a different input.

Although the form factor of smart glasses influenced the design of game inputs, there was almost no difference in the user preference ratings for user-defined game inputs between the 2 different glasses: ($F_{0.05}(1, 2446)$=.36, p=.549).

4.3 Classification of Game Inputs

4.3.1 Taxonomy of Game Input

As the authors, we manually classified the input actions along four dimensions: *form*, *binding*, *nature*, and *flow*. Within each dimension, there are multiple categories as shown in Table 4.3. To verify the objectivity (or inter-rater reliability), we invited an independent rater who performed the same categorization using 170 trials (10 trials were randomly selected for each task). The inter-rater reliability is shown in Table 4.2. The lowest Kappa

#	Task	Kappa Value
1	Select single from many	0.863
2	Vertical menu	1.000
3	Horizontal menu	0.688
4	Move left and right	0.825
5	Move in 4 directions	1.000
6	Switch 2 objects	0.804
7	Move object to position	1.000
8	Draw a path	1.000
9	Throw an object (in-2D)	1.000
10	Follow the beats	0.697
11	Rotate an object (Z-axis)	0.867
12	Rotate an object (Y-axis)	1.000
13	Avatar jump	0.867
14	Avatar 3D move	0.880
15	Avatar attack	1.000
16	Avatar squat	0.878
17	Control 3D viewport	0.878
	Average	**0.897**

Table 4.2: Inter-rater reliability for each task.

value, .688, is greater than .6, which is rated as *substantial* and thus is sufficient to establish the validity of the categorization. In addition, the average Kappa value is .897. A Kappa value of .8 and higher is considered *almost perfect*[23].

The scope of the *Form* dimension is applied separately to different interaction methods. There are 11 *Form* categories with *touch* input. 4 of them (*palm*, *back of hand*, *forearm*, *wrist*) are performed with both hands. 2 of them (*leg*, *face*) use a single hand to interact with other body parts. *fingers* is a single hand input and it is merely an interaction between fingers, e.g. a pinch. And rest of them (*ring*, *watch*, *glasses*, *necklace*) are interactions with accessories. There are 4 form categories with the *non-touch* interaction method. *Finger* is a special case of *hand*, but it is worth distinguishing because of its similarity to mouse actions and direct-touch.

In the *Nature* dimension, *symbolic* inputs are visual depiction. For example, a user poses the v-sign in the air in order to select menu option 2, or forms his hand as a gun to throw an object. *physical* inputs with the virtual object should be similar to the real world interaction with the physical object. *metaphorical* inputs occur when an input acts on, with, or like something else. For instance, users trace a finger in a circle to simulate

Taxonomy of Game Inputs		
Form ***(touch)***	*palm*	I.b. finger and palm.
	fingers	I.b. fingers.
	leg	I.b. finger and leg.
	back of hand	I.b. finger and back of hand.
	forearm	I.b. finger and forearm.
	face	I.b. finger and face.
	wrist	I.b. finger and wrist.
	ring	I.b. finger and ring.
	watch	I.b. finger and watch.
	glasses	I.b. finger and glasses.
	necklace	I.b. finger and necklace.
Form ***(non-touch)***	*finger*	Using finger to perform in-air gesture.
	hand	Using hand to perform in-air gesture.
	head	Using head to perform input.
	voice	Using voice control.
Binding	*direct*	Directly control in front of screen.
	surface	Absolute mapping screen to surface.
	independent	No binding b. screen and input.
Nature	*symbolic*	Input visually depicts a symbol.
	physical	Input acts physically on objects.
	metaphorical	Input indicates a metaphor.
	abstract	Input mapping is arbitrary.
Flow	*discrete*	Response occurs *after the user acts.*
	continuous	Response occurs *before the user acts.*

Table 4.3: Taxonomy of game inputs based on 2448 input actions. The abbreviation “I.b.” means “Interaction between”. The abbreviation “b.” means “between”.

the "object rotation", or view the palm as a trackpad to perform gestures. As it should be, the input itself is usually not enough to reveal its metaphorical nature; the answer lies in the user's mental model which could be understood by the interview afterwards. Finally, *abstract* inputs have no *symbolic*, *physical*, or *metaphorical* connection to their game tasks. The mapping is arbitrary, which does not necessarily mean it is poor. Pinch-touching thumb and index finger to perform "avatar jump", for example, would be an abstract input.

The *Binding* dimension is defined as the relationship between the input area and the smart glass's screen. *direct* binding means the user performs the inputs in the screen region directly, such as using an in-air gesture right in front of the screen to drag or touch virtual objects. A game input binding is called a *surface* binding if the user absolutely maps the screen onto another surface and performs the game inputs on it. Dragging a finger on the palm to move the object on the screen, for example, is a *surface* binding input. For inputs categorized as *independent* inputs it means that there is no binding between the screen and the input area. Thus the input can be performed in any position, like the "pinch to jump".

A game input's *Flow* is *discrete* if the input is performed, delimited, recognized, and responded to as an event. An example is punching in the air to perform "avatar attack". *Flow* is *continuous* if ongoing recognition is required, such as during most of our participants' "Control 3D viewport" rotating the imaginary camera by hands.

4.3.2 Taxonometric Breakdown of Input Actions in our Data

We found that our taxonomy adequately describes even widely differing input actions made by our users. Figure 4.1 and 4.2 show for each dimension the percentage distribution between the categories. The *Form* dimension for *touch* input and the the *Form* dimension for *non-touch* input are both dominated with hand related input. And *Form* dimension for *touch* input is consisted of *On-Body*(80.3%) and *Wearable*(19.7%) interaction. We found that the forms of *touch* inputs are more complicated than those of *non-touch*. Nonetheless, the binding of *touch* is more consistent. About 75% of the inputs are independent of the screen. In addition, we were surprised that no user designed a *direct* binding or *physical* nature input with *touch* input. And we found that for the *Flow* dimension, the percentage distribution is similar to *touch* and *non-touch* interaction.

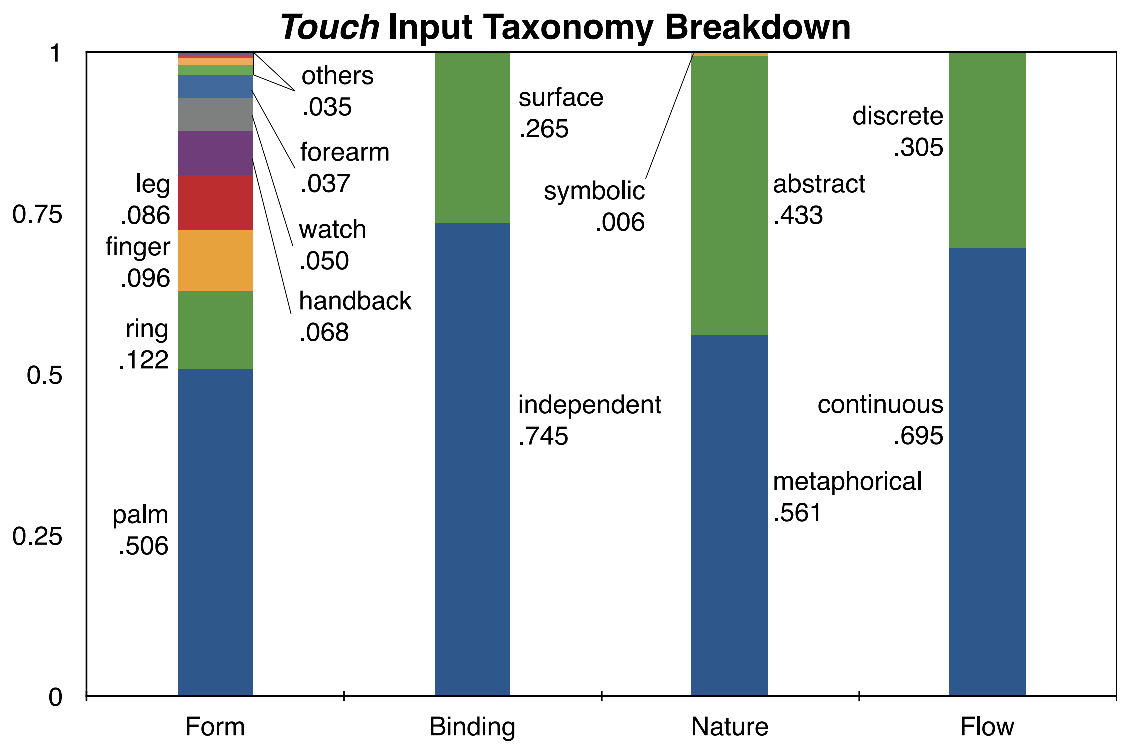

Figure 4.1: Percentage of game inputs in each taxonomy dimension with *touch* interaction. The "others" on the form dimension is consisted of glasses (0.0164), necklace (0.0088), face (0.0075) and wrist (0.0025).

4.4 User-Defined Game Input Set

The goal of this work is to present a user-defined game input set for smart glasses used in a public environment. This section gives the process by which the set was created and properties of the set. Unlike the input sets for existing games on smart glasses, the set we have found is based on observed user behavior.

4.4.1 Agreement

All 24 participants have provided game input for each and every game task, smart glasses form and interaction method. For each game task, we made groups of identical actions

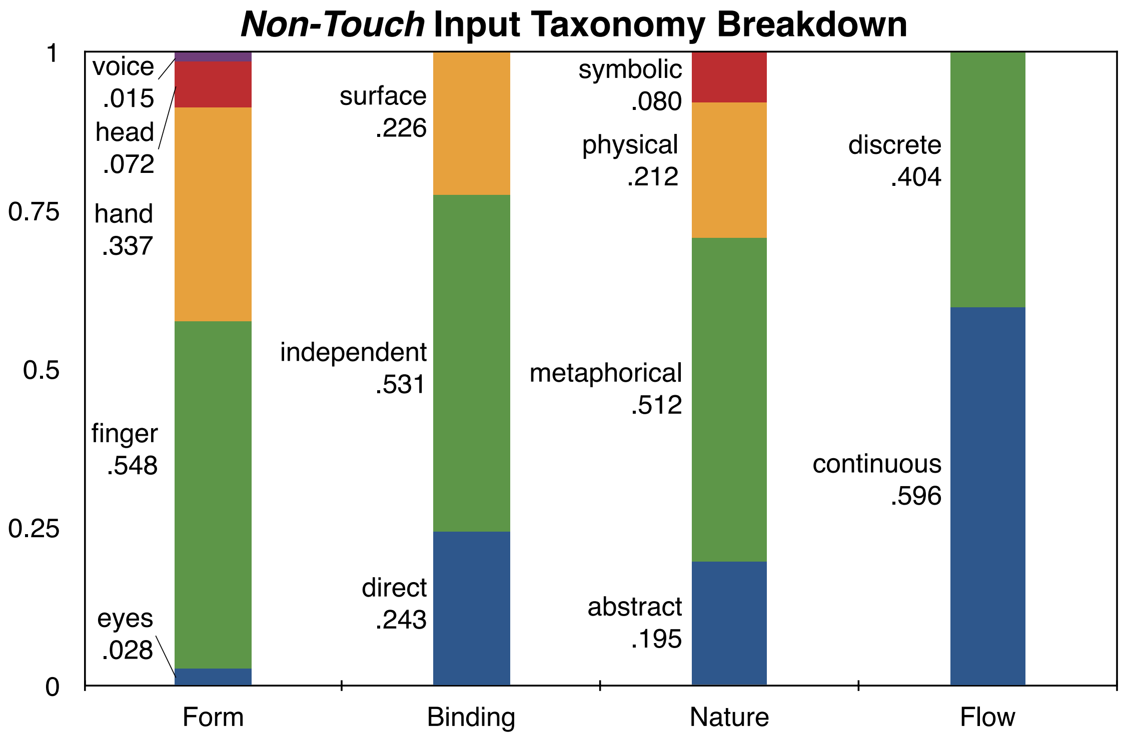

Figure 4.2: Percentage of game inputs in each taxonomy dimension with *non-touch* interaction.

used by participants. Group size was then used to compute an *Agreement Score* that reflects the consensus among participants regarding the action used for a certain game task. A task with a .31 agreement score means that, two randomly picked participants will have a 31% chance to perform an identical input action for this task. The definition and formula of the agreement score can be found in previous work. [42].

$$A = \frac{\sum_{t \epsilon T} \sum_{P_i \subseteq P_t} \left(\frac{|P_i|}{|P_t|} \right)^2}{|T|} \tag{4.1}$$

In eq. 1, t is a task in the set of all tasks T, P_t is the set of proposed input actions for task t, and P_i is a subset of identical input actions from P_t. The range for A is $[|P_t|^{-1}, 1]$. As an example, consider the agreement for *draw a path* with *touch* input, it had four groups of identical input actions with group sizes 34, 4, 5 and 5. we compute

$$A_{touch-path} = \left(\frac{34}{48}\right)^2 + \left(\frac{4}{48}\right)^2 + \left(\frac{5}{48}\right)^2 + \left(\frac{5}{48}\right)^2 = 0.53 \tag{4.2}$$

The participant agreement for our study is pictured in Figure 4.3. The overall agreement for *touch* and *non-touch* inputs were A_{touch}=0.25 and $A_{non-touch}$=0.27, respectively. When comparing the agreement of *touch* and *non-touch* inputs, we clearly see that their patterns are extremely similar. The average difference of agreement between these two interaction methods was .056. This implies that the agreement score was influenced more by the game tasks than by the interaction methods.

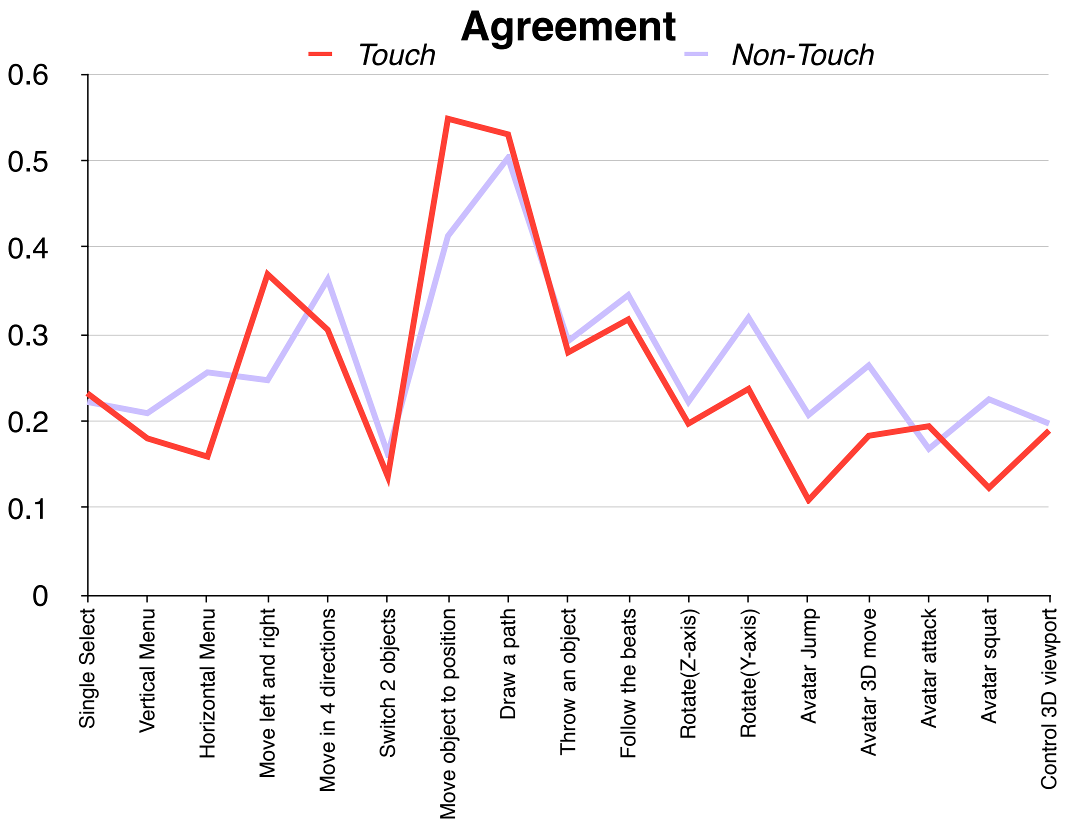

Figure 4.3: Agreement for each game task. The tasks are listed in the same order as they appear in Table 3.1.

4.4.2 Properties of the User-defined Game Input Set

The user-defined game input set was developed by taking the largest groups with identical input actions for each game task and assigning those actions to the game tasks. The resulting user-defined game input set covers 41.32% and 40.07% of all game inputs proposed for both the *touch* and the *non-touch* interaction class respectively. Our user defined set is useful, therefore, not just for what it contains, but also for what it omits.

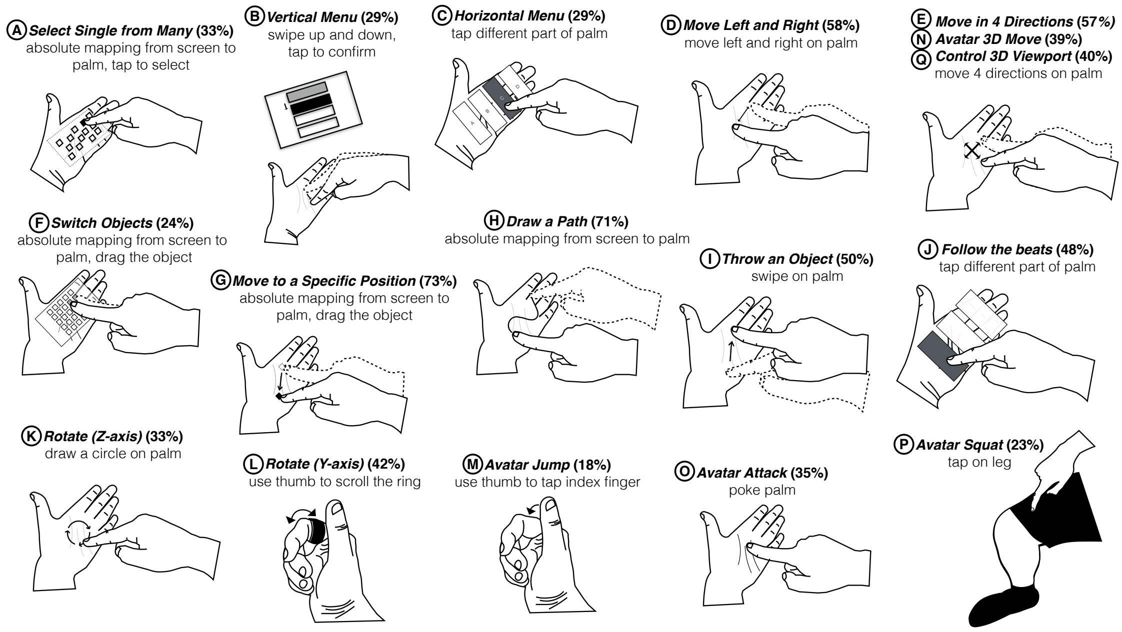

Figure 4.4: The user-defined game input set with *touch* inputs. The percentages indicate the portion of users who performed the pictured input action for the game task. Note that, there are 3 tasks ("Move in 4 directions", "Avatar 3D Move", and "Control 3D viewport") have been assigned with an identical input action.

Figure 4.5: The user-defined game input set with *non-touch* interaction, The percentages indicate the portion of users who performed the pictured input action for the game task.

All the inputs in our final *touch* input set are finger based (see Figure 4.4). Most of them use a single finger tip to perform gestures on different surfaces. The most preferable surface for *touch* input is the hand palm. This way, the hand palm acts as a trackpad or a proxy touch-screen. The other touch inputs are finger interactions with single hand. More specifically, participants used their thumb to interact with their index finger or the ring on the finger.

For *non-touch* input set (see Figure 4.5), even though we informed users beforehand that they were not limited to using their hands when providing game input, the results show that users still preferred to use hand input over voice control, eye gestures and head tilting. Additionally, users would make use of direct-control if they had to perform precise tasks, such as selecting an object from many or moving an object to a specific position. On the other hand, for tasks with lower precision requirements, such as selecting a single option from 4 or making an avatar jump, users would prefer using an indirect-control. For examples: the user taps 4 different areas in front of their chest or the user raises his hands slightly.

4.4.3 Taxonometric Breakdown of User-Defined Game Inputs

As expected, the taxonometric breakdown of the final user-defined game input set (Figure 4.4 and 4.5) is similar to the proportions of all control actions proposed (Figure 4.1 and 4.2). Across all taxonomy categories, the average difference between these two sets was only 5.61%, (*touch* input 6.31% and *non-touch* input 4.91%, respectively).

4.5 Mental Model Observations

4.5.1 Social Acceptance and Input Area

To our surprise, approximately 63% of the in-air gestures were not performed in front of the face (See Figure 4.6.2). This behavior conflicts with the current "Google Glass" design. There were 7 participants who performed most gestures in front of their face. They indicated that input in front of the face was more precise and intuitive. At the same time, the other 17 participants preferred to perform in-air gestures in front of or below their chest. Among them, there were 3 participants who didn't perform a single in-air gesture in front of their face. These users indicated that moving a finger in front of their face was weird and not socially acceptable. They also noted that there was a hand fatigue problem if they had to lift their hand in front of their face all the time, so they thought that it was not suitable for gaming.

4.5.2 Bias by Existing Game Input

Although we were careful not to show elements from traditional game platforms like PCs, consoles and mobile games, participants still often reasoned based on their previous gaming experience. For example, some input actions were performed as if using a touch-screen in front of their face (see Figure 4.5{A,F,G}). Some actions were like using an imaginary trackpad on an in-air surface or on the hand palm(see Figure 4.4{B,D,E,H,I,K,N,Q} and Figure 4.5{H,I}). Even with simple shapes and basic characters, it was clear how deeply rooted the previous gaming experience is. Some quotes reveal this: "So I just click a button like on a game controller", "Can I just imagine there is a trackpad on my palm?" and "It's an imaginary touch-screen."

4.5.3 Identical Gestures on Different Surfaces

In our study, we found several identical gestures performed by our users on different surfaces. Take the task "Move in 4 directions" for example, although 57% of the gestures were performed by moving a finger on the palm with *touch* inputs(Figure 4.4.E). The rest of the gestures were mostly using identical gestures (moving a finger), but then on the different surfaces such as the back of the hand, the leg, the forearm and on the even face. The same phenomenon could also be observed when comparing the game input of the user-defined input set with *touch* and *non-touch* inputs for the tasks "Move left and right", "Move in 4 directions", "Draw a Path", "Throw an Object". Participants used the hand palm or an imaginary in-air surface. (See Figure 4.4 and Figure 4.5 {D,E,H,I}).

In these cases, the surface did not influence the meaning of the gestures. We have asked users why they chose the palm as their input area. The general response was that it required the least physical movement, such as "I chose the left palm to perform a gesture on because it is near to my right hand".

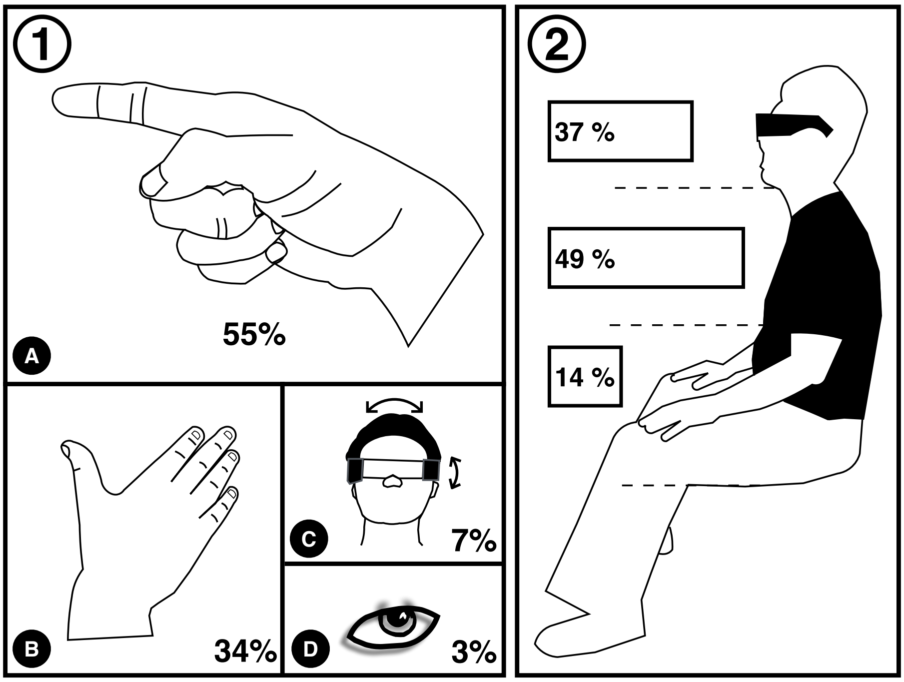

Figure 4.6: 1.The top 4 *non-touch* interaction forms. Percentages indicate the portion of *non-touch* game inputs that consisted of the pictured input. (A)Using a finger to perform an in-air gesture. (B)Using the full hand to perform an in-air gesture. (C)Using head-tilting to perform game input. (D)Using eye-gestures to perform game input. 2.The distribution of the in-air gesture input area. Half of the in-air gestures (49%) were performed in front of the chest, 14% in front of or below the belly, and only 37% of the gestures were performed in front of the face.

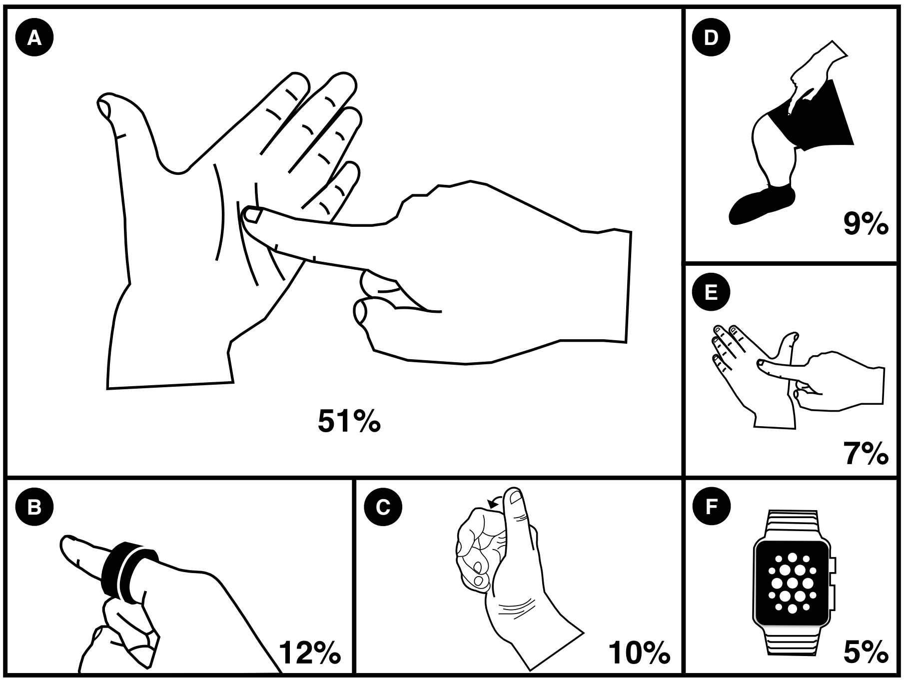

Figure 4.7: The top 6 *touch* input forms. Percentage indicates the portion of *touch* input actions. (A)Interaction between finger and palm. (B)Interaction between finger and ring. (C)Interaction between fingers. (D)Interaction between finger and leg. (E)Interaction between finger and back of hand. (F)Interaction between finger and watch.

Chapter 5

Discussion

5.1 Implications for Touch Input Technology

Our results showed that the hand palm was the most favorite area for users to perform *touch* inputs on. Half of the game inputs with *touch* input used a finger to perform a gesture on the palm (See Figure 4.7). According to the mental model mentioned above, users utilized the metaphor of a trackpad and a touch-screen on the palm in several cases. Since this metaphor leads to the same input actions as on trackpads and touch-screens, current gesture interpreting algorithms like Dollar N[10] could be employed here.

5.2 Implications for Non-Touch Interaction Technology

For *non-touch* interaction methods, our taxonomy shows that performing in-air gestures with fingers and hands are still the dominant forms for smart glass gaming (Figure 4.6.1{A,B}). There was only a small number of participants that used head-gestures, eye-gestures or voice controls, 7%, 3% and 1% respectively. Before our study, both Google Glass and Mime[15, 9] supplied their own in-air gesture sets to increase the diversity of their input. However, our results show that 63% of the in-air gestures are not performed in front of the user's face in the public space due to the social acceptance issues and physical tiring problems mentioned before (see Figure 4.6.2). Therefore, if the developers of head-worn devices want to implement in-air gestures for input, they will need to have the capability to sense gestures in a wide range of areas near the user other than only right in front of the face. Take CV-based sensing technologies for example, instead of a regular lens, we could use a wide-angle lens or fish-eye lens to implement a system to cater to the user's preference [6].

5.3 Implications for Game Design

According to the agreement scores we found that, no matter if using *touch* interaction or *non-touch* input, the average agreement between users was only .26, and the highest agreement was just about .55 (see Figure 4.3). In this case, guessing the game inputs would become a frustrating experience for players. It indicated that game developers should design the visual guide carefully to lead users performing the input action, or show an instruction to explain the input methods.

5.4 Contribution to Non-gaming Scenarios

When we first set out to explore interaction design for smart glasses, we focused on a specific domain in order to gain deeper insight and to keep the study tractable. Looking back at the results, some the study findings do apply to non-gaming scenarios. For example, many of our tasks are also used in non-gaming applications, such as “Select single from many”, “Vertical menu”, “Horizontal menu”, “Move left and right”, “Move in 4 directions”, “Move object to position”, “Draw a path” and “Rotate an object”. Also, study results showed several facts that are useful for general cases: (1) Social acceptance of input is a significant concern in public space; (2) Performing in-air gestures in front of the face is weird and not socially acceptable; (3) If the input surface does not have an influence on the meaning of the gestures, users prefer to perform the gestures on a surface reached with least movement.

5.5 Limitation and Next Steps

As we know, there are many different places known as public space, and users may behave differently in each specific place. Furthermore, in our study, we did not ask users to define any input actions to interact with tangible objects in public space, such as, tables or chairs in the cafe shop. We only made participants experience two types of smart glasses. Therefore, our user-defined game input set might not be suitable to be applied to games on other types of head-worn devices. Moreover, our participants were all literate Taiwanese adults; undoubtedly, children, elders, participants from other cultures, or uneducated participants would behave differently. That is to say, these issues are worthy of investigation, but exceed the range of our current work.

An important next step is to validate our user-defined game input set with a wearable system, which can sense all *touch* and *non-touch* input actions listed in our set.

Chapter 6

Conclusion

This paper explored user-defined game input for smart glasses beyond the capabilities of current sensors, and focused on gaming interaction in a public setting. We conducted a user-defined input study with 24 participants, each performing 17 common game control tasks using *handheld*, *touch* and *non-touch* interaction methods with two form factors of smart glasses in a public cafe, which lead to a total of 2448 game inputs. Our results indicate that participants significantly preferred *non-touch* interactions over *handheld* interactions (3.81 vs 3.68, p <0.01). And the most frequently used body surface was the palm (51%). Also, participants preferred using in-air gestures in front of the torso over gestures in front of the face(63% vs 37%) due to concerns with social acceptance and the hand fatigue. Furthermore, we indicated the mismatch between participants' preferred input methods and those supported by current smart glasses. Finally, we presented insight into users' mental models, and an understanding of implications for input technology and interface design. This work represents a necessary step in bringing glasses gaming closer to the hands and minds of smart glasses users.

Bibliography

[1] R. Aigner, D. Wigdor, H. Benko, M. Haller, D. Lindbauer, A. Ion, S. Zhao, and J. Koh. Understanding mid-air hand gestures: A study of human preferences in usage of gesture types for hci. *Microsoft Research TechReport MSR-TR-2012-111*, 2012.

[2] L. Anthony, Q. Brown, J. Nias, B. Tate, and S. Mohan. Interaction and recognition challenges in interpreting children's touch and gesture input on mobile devices. ITS'12, pages 225–234. ACM.

[3] App Annie. http://www.appannie.com.

[4] T. Baba, T. Ushiama, R. Tsuruno, and K. Tomimatsu. Video game that uses skin contact as controller input. SIGGRAPH'07. ACM.

[5] A. Biskupski, A. R. Fender, T. M. Feuchtner, M. Karsten, and J. D. Willaredt. Drunken ed: A balance game for public large screen displays. CHI EA'14, pages 289–292. ACM.

[6] L. Chan, C.-H. Hsieh, Y.-L. Chen, S. Yang, D.-Y. Huang, R.-H. Liang, and B.-Y. Chen. Cyclops: Wearable and single-piece full-body gesture input device. CHI'15. ACM.

[7] L. Chan, R.-H. Liang, M.-C. Tsai, K.-Y. Cheng, C.-H. Su, M. Y. Chen, W.-H. Cheng, and B.-Y. Chen. Fingerpad: Private and subtle interaction using fingertips. UIST'13. ACM.

[8] S. Christian, J. Alves, A. Ferreira, D. Jesus, R. Freitas, and N. Vieira. Volcano salvation: Interaction through gesture and head tracking. CHI EA'14, pages 297–300. ACM.

[9] A. Colaço, A. Kirmani, H. S. Yang, N.-W. Gong, C. Schmandt, and V. K. Goyal. Mime: Compact, low power 3d gesture sensing for interaction with head mounted displays. UIST'13, pages 227–236. ACM.

[10] Dollar N Multistroke Recognizer. http://depts.washington.edu/aimgroup/proj/dollar/ndollar.html.

[11] J. Epps, S. Lichman, and M. Wu. A study of hand shape use in tabletop gesture interaction. CHI'06, pages 748–753.

[12] Epson BT-100 Specs. http://www.epson.com.au/products/ProjectorAccessories/Moverio_BT-100_specs.asp.

[13] Essential Facts About The Computer And Video Game Industry. http://www.theesa.com/facts/pdfs/esa_ef_2014.pdf.

[14] GameStop. http://www.gamestop.com.

[15] Google Glass wiki. http://en.wikipedia.org/wiki/Google_Glass#Features.

[16] D. Grijincu, M. A. Nacenta, and P. O. Kristensson. User-defined interface gestures: Dataset and analysis. ITS'14, pages 25–34. ACM.

[17] S. Gustafson, C. Holz, and P. Baudisch. Imaginary phone: Learning imaginary interfaces by transferring spatial memory from a familiar device. UIST'11, pages 283–292. ACM.

[18] S. Harada, J. O. Wobbrock, and J. A. Landay. Voice games: Investigation into the use of non-speech voice input for making computer games more accessible. INTERACT'11, pages 11–29. Springer-Verlag.

[19] C. Harrison, H. Benko, and A. D. Wilson. Omnitouch: Wearable multitouch interaction everywhere. UIST'11, pages 441–450. ACM.

[20] C. Harrison, D. Tan, and D. Morris. Skinput: Appropriating the body as an input surface. CHI'10, pages 453–462. ACM.

[21] C.-Y. Hsu, Y.-C. Tung, H.-Y. Wang, S. Chyou, J.-W. Lin, and M. Y. Chen. Glass shooter: Exploring first-person shooter game control with google glass. ICMI'14, pages 70–71. ACM.

[22] L. Jing, Z. Cheng, Y. Zhou, J. Wang, and T. Huang. Magic ring: A self-contained gesture input device on finger. MUM'13, pages 39:1–39:4. ACM.

[23] Cohen's kappa - Wikipedia, the free encyclopedia. http://en.wikipedia.org/wiki/Cohen's_kappa.

[24] M. Karam et al. A taxonomy of gestures in human computer interactions. 2005.

[25] D. Kim, O. Hilliges, S. Izadi, A. D. Butler, J. Chen, I. Oikonomidis, and P. Olivier. Digits: Freehand 3d interactions anywhere using a wrist-worn gloveless sensor. UIST'12, pages 167–176. ACM.

[26] H.-N. Liang, C. Williams, M. Semegen, W. Stuerzlinger, and P. Irani. User-defined surface+motion gestures for 3d manipulation of objects at a distance through a mobile device. APCHI'12, pages 299–308. ACM.

[27] Google Glass Mini Games. https://developers.google.com/glass/samples/mini-games.

[28] C. S. Montero, J. Alexander, M. T. Marshall, and S. Subramanian. Would you do that?: Understanding social acceptance of gestural interfaces. MobileHCI'10, pages 275–278. ACM.

[29] M. R. Morris. Web on the wall: Insights from a multimodal interaction elicitation study. ITS'12, pages 95–104. ACM.

[30] L. E. Nacke, M. Kalyn, C. Lough, and R. L. Mandryk. Biofeedback game design: Using direct and indirect physiological control to enhance game interaction. CHI'11, pages 103–112. ACM.

[31] M. Nielsen, M. Störring, T. B. Moeslund, and E. Granum. A procedure for developing intuitive and ergonomic gesture interfaces for hci. In *Gesture-Based Communication in Human-Computer Interaction*, pages 409–420. Springer, 2004.

[32] T. Piumsomboon, A. Clark, M. Billinghurst, and A. Cockburn. User-defined gestures for augmented reality. CHI'13, pages 955–960, New York, NY, USA. ACM.

[33] D. Pyryeskin, M. Hancock, and J. Hoey. Comparing elicited gestures to designer-created gestures for selection above a multitouch surface. ITS'12, pages 1–10. ACM.

[34] S. Reis. Expanding the magic circle in pervasive casual play. ICEC'12, pages 486–489. Springer-Verlag.

[35] M. Serrano, B. M. Ens, and P. P. Irani. Exploring the use of hand-to-face input for interacting with head-worn displays. CHI'14, pages 3181–3190. ACM.

[36] A. J. Sporka, S. H. Kurniawan, M. Mahmud, and P. Slavík. Non-speech input and speech recognition for real-time control of computer games. Assets'06, pages 213–220. ACM.

[37] Steam. http://store.steampowered.com.

[38] Top 90 Casual Games List. http://ppt.cc/hO2j, Analyzed at 2014-08-14.

[39] VGChartz. http://www.vgchartz.com.

[40] S. Vickers, H. Istance, and A. Hyrskykari. Performing locomotion tasks in immersive computer games with an adapted eye-tracking interface. *ACM Trans. Access. Comput.*, 5(1):2:1–2:33, Sept. 2013.

[41] J. R. Williamson, S. Brewster, and R. Vennelakanti. Mo!games: Evaluating mobile gestures in the wild. ICMI'13, pages 173–180. ACM.

[42] J. O. Wobbrock, H. H. Aung, B. Rothrock, and B. A. Myers. Maximizing the guessability of symbolic input. CHI EA'05, pages 1869–1872. ACM.

[43] J. O. Wobbrock, M. R. Morris, and A. D. Wilson. User-defined gestures for surface computing. CHI'09, pages 1083–1092. ACM.