

User-Defined Motion Gestures for Mobile Interaction

Jaime Ruiz*

University of Waterloo
Waterloo, ON, Canada
jgruiz@cs.uwaterloo.ca

Yang Li

Google Research
Mountain View, CA, USA
yangli@acm.org

Edward Lank

University of Waterloo
Waterloo, ON, Canada
lank@cs.uwaterloo.ca

ABSTRACT

Modern smartphones contain sophisticated sensors to monitor three-dimensional movement of the device. These sensors permit devices to recognize motion gestures—deliberate movements of the device by end-users to invoke commands. However, little is known about best-practices in motion gesture design for the mobile computing paradigm. To address this issue, we present the results of a guessability study that elicits end-user motion gestures to invoke commands on a smartphone device. We demonstrate that consensus exists among our participants on parameters of movement and on mappings of motion gestures onto commands. We use this consensus to develop a taxonomy for motion gestures and to specify an end-user inspired motion gesture set. We highlight the implications of this work to the design of smartphone applications and hardware. Finally, we argue that our results influence best practices in design for all gestural interfaces.

Author Keywords

Motion gestures, sensors, mobile interaction.

ACM Classification Keywords

H5.2. User Interfaces--Input devices and strategies.

General Terms

Design

INTRODUCTION

While smartphones combine several tasks (e.g., voice and data communication, multimedia consumption, mobile gaming, and GPS navigation) into one package, their form factor is also limiting in both input and output. To allow the device to fit into a pocket or purse, screens are small and keyboards are thumb-sized. On many devices the thumb keyboard has been replaced by a soft-keyboard displayed on the screen to minimize the size and weight of the device.

Two primary input modalities are commonly supported by soft-keyboard based smartphones: The first is a touchscreen display; and the second is a set of motion

sensors: accelerometers, gyroscopes, orientation sensors (vs. gravity), etc. The two inputs recognized by these devices are different types of gestures. Users can gesture on the device in two dimensions, using the touchscreen of the smartphone as a mobile surface computer. We call these two-dimensional gestures *surface gestures*. Users can also gesture with the device, in three dimensions, by translating or rotating the device. We call these three-dimensional gestures *motion gestures*.

In this research, we focus specifically on *motion gestures*. Researchers have proposed the use of motion gestures for a variety of input tasks: for example, to navigate maps or images [18], to input text [11,17,26], to control a cursor [25], and to verify user identity [12].

Recently, Wobbrock et al. [28] addressed the lack of understanding of the design space for surface gestures. However, many similar questions about motion gesture design are also unanswered by past research. What parameters do users manipulate to create different motion gestures (i.e. differences in path, in kinematics, etc.)? Is there a “design-space” or taxonomy of the different dimensions that designers can manipulate in the creation of these gestures? Is there an end-user consensus set of user-specified motion gestures that eliminates the need for designers to arbitrarily create their own motion gestures? Finally, is there a logical mapping of motion gestures onto device commands?

In this paper, we describe the results of a *guessability study* [27] for motion gestures which elicits natural gestures from end-users as follows: given a task to perform with the device (e.g. answer the phone, navigate East in a map), participants were asked to specify a motion gesture that would execute that task.

The results of the study yield two specific research contributions to motion gesture design. First, when participants were asked to specify motion gestures for many common smartphone effects including answering the phone, ignoring a call, or navigating within applications, there was broad unscripted agreement on the gestures. As a result, we can specify an end-user motion gesture set for many common smartphone tasks, analogous to Wobbrock et al.’s end-user surface gesture set [28]. Second, we use

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.

CHI 2011, May 7–12, 2011, Vancouver, BC, Canada.

Copyright 2011 ACM 978-1-4503-0267-8/11/05...\$10.00.

*This work was conducted during an internship at Google Research.

measurements from both the input sensors of the smartphone and video recordings of motion gestures created by participants to specify a taxonomy of the parameters that can be manipulated to differentiate between different motion gestures. This taxonomy represents the design space of motion gestures.

The implications of this research to the design of smartphone appliances are two-fold. First, from the perspective of smartphone application designers, the taxonomy of physical gestures and our understanding of agreement for user-defined gestures allow the creation of a more natural set of user gestures. They also allow a more effective mapping of motion gestures onto commands invoked on the system. Second, from the perspective of companies that create smartphones and smartphone operating systems, this study provides guidance in the design of sensors (i.e. what features of three dimensional motion must we distinguish between) and toolkits (i.e. what gestures should be recognized and accessible to application context) to support motion gesture interaction at both the application and the system level.

More broadly, the results reported in this paper significantly extend our understanding of gestural interaction both in two dimensions (surface gestures) and in three dimensions (motion gestures). Broad agreement exists among users on gesture sets, both in two [28] and six degrees of freedom. As well, users' past experiences with desktop computers or with smartphones inform a logical mapping of causes onto effects for both surface and mobile computing. Our work suggests that these consistent logical mappings would extend to paradigms beyond just surface and mobile computing. If the effects persist for other computing paradigms, then the design task for gestural interaction becomes one of manipulating established taxonomies of gestures while preserving the logical mappings of causes to effects as specified by end users. The downside of this is that the constraints on gesture designers increase, i.e. the logical mapping must be preserved. The benefit is that, whether in 2D or 3D, natural gestures and natural mappings can potentially be identified by conducting a basic guessability study with prospective users.

The rest of this paper is organized as follows. We first explore related work in user-specified gesture sets and in physical gestures for device control. Next, we describe our study methodology, including our participants and the set of tasks that we examine. We describe the qualitative data and the taxonomy, the specific results of our observational study. Finally, we discuss in more detail the broader implications of this work.

RELATED WORK

A majority of the prior work on classifying human gestures has focused on human discourse (see [28] for a review). In this section, we focus on work which explores the classification of human gestures in relation to the dialog between a human user and an interactive device. Within this

space, research exists in both surface gestures and motion gestures. We also describe techniques for elicitation studies as a motivation for our research approach to motion gestures.

Surface Gesture Research

In the domain of surface computing, surface gestures have been used by groups of users to support cooperative work with systems and by single users to issue commands to the system. Tang [23] observed that gestures play an important role in communicating significant information for small groups around drawing interfaces. He observed that gestures are used to express ideas, demonstrate a sequence of actions, and mediate group interaction. Morris et al. [15] described a classification, or design space, for collaborative gestures resulting from their evaluation of a system for cooperative art and photo manipulation. Their classification identified seven design axes relevant to cooperative gesture interaction: symmetry, parallelism, proxemic distance, additivity, identity-awareness, number of users, and number of devices.

In work examining gestures for single-user interaction, Wobbrock et al. [28] present a taxonomy of surface gestures based on user behavior. Based on a collection of gestures from twenty participants, their taxonomy classifies gestures into four dimensions: form, nature, binding, and flow. They also create a user-specified gesture set. More recently, they evaluated this gesture set against a gesture set created by a designer and showed that the user-specified gesture set is easier for users to master [16]. Wobbrock et al.'s work in surface gestures is a strong justification for elicitation studies in the design of gesture sets.

Motion Gesture Research

To our knowledge no research has been published describing the classification of motion gestures. As well, little research has been done on end-user elicitation of motion gestures. Research on motion gestures has been focused on interaction techniques using motion input and tools to design motion gestures.

Rekimoto [18] was credited for proposing one of the earliest systems to use motion input to interact with virtual objects. Rekimoto demonstrated how mapping motion to tilt can be used for selecting menu items, interacting with scroll bars, panning or zooming around a digital workspace, and performing complex tasks such as 3D object manipulations.

Harrison et al. [6], Small & Ishii [22], and Bartlett [4] extended the use of tilt sensors to enable navigating through widgets on mobile devices. Hinckley et al. [9] proposed using tilt on a mobile device to allow a user to change screen orientation—a feature now commonly found on many devices. Motion input has also been used for text input [11,17,26], controlling a cursor [25], and user verification [12].

Research efforts in physical gestures have also been targeted at designers of systems that use physical gestures. Exemplar [8] allows quick motion gesture design using demonstration and direct manipulation to edit a gesture. MAGIC [3] allows designers to design motion gestures by demonstration, and incorporates tools that provide information about performance. MAGIC also allows designers to test for false positives, internal consistency, and distinguishability between classes of gestures to improve the recognition rate of motion gestures created by designers.

What little research exists on end-user elicitation of motion gestures has been done in support of multimodal interaction. In this domain, Mignot et al. [13] studied the use of speech and gesture and found that gestures were used for simple and direct commands, while speech was more commonly used for abstract commands. In their work on augmented reality offices, Volda et al. [24] asked participants to create gesture and/or voice commands for accessing objects on multiple displays. They found that people overwhelmingly used finger pointing. While some elicitation of motion gestures exists in the multimodal interaction community, the work has typically explored physical gesture input as an add-on to voice-based commands. Research on the use of motion gestures as a stand-alone input modality has not been explored by these researchers.

Conducting Elicitation Studies

Eliciting input from users is a common practice and is the basis for *participatory design* [21]. Our approach of prompting users with the effects of an action and having them perform a gesture has been used to develop a command line email interface [5], unistroke gestures [27], and gestures for surface computing [28].

DEVELOPING A USER-DEFINED GESTURE SET

To explore user-defined gestures, we elicited input from 20 participants. Participants were asked to design and perform a motion gesture with a smartphone device (a cause) that could be used to execute a task on the smartphone (an effect). Nineteen tasks were presented to the participants during the study (Table 1). Participants used the think-aloud protocol and supplied subjective preference ratings for each gesture.

As the goal of the study was to elicit a set of end-user gestures, we did not want participants to focus on recognizer issues or current smartphone sensing technology. As a result, no recognizer feedback was provided to participants during performance of the gestures. We also encouraged the participants to ignore recognition issues by instructing them to treat the smartphone device as a “magic brick” capable of understanding and recognizing any gesture they might wish to perform. Our rationale for these decisions was the same as the rationale expressed in Wobbrock et al.’s surface gesture work [28]. Specifically, we wished to remove the *gulf of execution* [10] from the dialog between the user and the device, i.e. to observe the

Category	Sub-Category	Task Name
Action	System/Phone	Answer Call
		Hang-up Call
		Ignore Call
		Voice Search
		Place Call
	Application	Act on Selection
Navigation	System/Phone	Home Screen
		App switch Next
		App switch Previous
	Application	Next (Vertical)
		Previous (Vertical)
		Next (Horizontal)
		Previous (Horizontal)
		Pan Left
		Pan Right
		Pan Up
		Pan Down
		Zoom In
		Zoom Out

Table 1. The list of tasks presented to participants grouped by category.

users’ unrevised behavior without users being influenced by the ability of the system to recognize gestures.

Each participant performed gestures for each of the tasks indicated in Table 1. The session was video recorded and custom software running on the phone recorded the data stream generated from the accelerometer. Each session took approximately one hour to complete. For each participant, a transcript of the recorded video was created to extract individual quotes and classify and label each motion gesture designed by the participant. The quotes were then clustered to identify common themes using a bottom-up, inductive analysis approach.

Selection of Tasks

Inclusion of tasks in our study was determined by first classifying tasks into two categories: *actions* and *navigation-based* tasks. Within these categories, we created two sub-categories: a task can either act on the system/phone (e.g. answering a phone call or switching to a previous application) or on a particular application (e.g. navigating a map in a GPS application, selecting text on the display). After grouping the tasks into these four sub-categories, a scenario representing each task was chosen for inclusion in the study. This method allowed us to create tasks that would be representative of the tasks used in a smartphone while minimizing duplication of tasks resulting

from application specific gestures. Categories and sub-categories for the tasks are shown in Table 1.

Participants

Grasping the concept of moving the device to invoke commands and clearly understanding the potential tasks that could be performed on a smartphone (both necessary to collect usable data in a guessability study) arguably require that the user have some experience with the device. Therefore, we intentionally recruited participants who indicated that they used a smartphone as their primary mobile device.

Twenty volunteers, ten males and ten females, between the ages of 21-44 (mean = 28, SD = 5.4) participated in the study. The participants all worked for a high-tech company but did not all hold technical positions. The volunteers received a \$30 gift certificate to an online bookseller for their participation.

Apparatus

Gestures were recorded using custom software developed using the Android SDK [1] for a Google Nexus One smartphone running Android 2.1. The software was responsible for logging the data stream of the accelerometer sensor and locking the screen to ensure no feedback was displayed to the participant. Additional software written in Java ran on the researcher's laptop and was responsible for recording the beginning and end of a gesture as well as the participant's subjective ratings.

Procedure

At the beginning of each experimental session, the researcher described the study to the participant and handed the participant the smartphone running the custom software. The 19 tasks were grouped into six sets of similar tasks. For example, one task set included effects that represented normal use of the phone: answering a call, muting the phone, ending a call. Another task set involved map navigation tasks such as panning and zooming.

For each set of tasks, the participant was presented with a sheet describing the overall set of tasks they were to invoke and listing each task in the group. Where appropriate (e.g., navigating Google Maps) a screenshot of an application was provided. Participants were instructed to read the information sheet and, in a talk-aloud method, design a motion gesture to represent each one of the listed tasks. Participants did not need to commit to a gesture until all gestures in the task set were designed to encourage participants to design a cohesive set of gestures for the set of tasks.

After designing the set of motion gestures for the given task set, the researcher asked the participant to perform each gesture five times on cue and then rate the gesture using a seven-point Likert scale on the following criteria:

- The gesture I picked is a good match for its intended use.
- The gesture I picked is easy to perform.

In addition, the user was also asked how often they would use the motion gesture if the gesture existed on a six-point scale ranging from never to very frequently.

The interview concluded with the interviewer asking the participants if they had suggestions of other tasks where motion gestures would be beneficial. Participants were then asked to design a gesture for each task they suggested. The purpose of this exercise was to assess if our proposed tasks had enough coverage of possible uses of the phone.

RESULTS

The data collected during our study included transcripts, the video recording, a set of gestures designed by our participants, subjective ratings of the set of gestures, and the data stream collected from the sensors while participants performed their gestures on the smartphone. From this data we present themes emerging from our interviews, a taxonomy for motion gestures, and a user-defined motion gesture set for mobile interaction.

Designing Motion Gestures

Transcripts of the recorded interviews were used to identify common themes that emerged from our study. The themes—which provide user-defined design heuristics for motion gestures—include mimicking normal use, applying real world metaphors, natural and consistent mappings, and providing feedback. We discuss each of these themes below.

Mimic Normal Use

Volunteers who designed gestures that mimicked motions occurring during normal use of the phone often perceived their gesture as being both a better fit to the task and easier to perform. In addition, there was a consensus among participants on the form of these gestures. This is especially evident in the design of a motion gesture to answer a call. For this task, 17 out of 20 users designed a gesture where users placed the phone to their ear. When asked to describe why they chose that gesture, participants often made comments describing the gesture as “natural”:

The first motion I would be doing is picking it up [and] bringing it to my ear...The most natural thing for me would be bringing it to my ear.[P16].

Real-world Metaphors

When participants were able to relate interacting with the mobile phone to interacting with a physical object, the gesture they designed consistently mimicked the use of a non-smartphone object. For example, to end a call, a majority of participants suggested removing the phone from the ear and turning the display face down parallel to the ground. When asked why they choose that gesture to represent the task, several participants noted that it mimicked the action of hanging up a phone receiver on an “old-fashioned” telephone.

Real-world metaphors do not always need to correspond directly to the phone. For example, when given the task of navigating to the Home Screen, half of the users selected

shaking the phone as the gesture. Users viewed navigating to home as “clearing what you are doing” [P6]. Users related clearing the contents of the phone to the action of clearing the contents of an Etch A Sketch [2]:

Why shaking? It’s almost like the Etch A Sketch where, when you want to start over, you shake it. [P20].

Natural and Consistent Mappings

Motion gestures differ from surface gestures in that the user interacts by using the device itself instead of interacting on the device with a finger or hardware button. To allow designers to create more intuitive motion gesture sets, it is important to understand the user’s mental model of how motion gestures map to the interaction of the device instead of relying on current mappings.

Tasks that were considered to be opposites of each other always resulted in a similar gesture but performed in the opposite direction, regardless of the proposed gesture. For example, a flick to the right was the most common gesture for next and a flick to the left was used by these same participants for previous.

Several sets of tasks were designed as navigational or scrolling tasks with the intention of determining the participant’s mental model of navigation (i.e., Is the participant controlling the viewport or the content?). Current touch interfaces often require the user to interact with the content while the viewport remains static. In contrast, when interacting with a scroll bar on a desktop PC the scroll bar controls the viewport.

Results from our study show that the preference of a participant depends on the plane in which she is interacting. In cases where the participant was interacting on the XY plane, i.e. navigating on a map, the consensus was that the interaction using motions should alter the viewport. In other words, to move to the left in a map, participants would move the phone to the left, similar to the interaction found in the peephole system [30]. Even those participants who first described interaction with the content performed gestures that required the viewport to move. For example, when asked to pan a map to the east (right), participants performed a gesture to the right, indicating the viewport would move to show the content east of the current position. When the interviewer mentioned this discrepancy between the description and the gesture, one participant responded:

I didn’t even notice I was doing it. [P15]

While moving in the XY plane resulted in viewport manipulations, when choosing gestures to zoom in and out of a map, i.e. interacting in the Z plane, the consensus was to perform a gesture to “move” the map closer to the participant’s face to zoom in and to move the map away from the participant’s face to zooming out. Therefore, instead of treating the phone as viewport, participants instead reverted to a real-world metaphor: a magnifying glass. Understanding the subtleties that exist between

mappings and metaphors is one challenge of implementing gesture set behaviors.

While map navigation used the metaphor of a viewport, the on-screen context had an effect on participants’ mappings. As part of the study we included two presentations of lists to determine if list orientation (horizontal or vertical) influenced the design decisions of our participants, and to determine whether list navigation and map navigation were analogous tasks. A majority of the participants shared the sentiments of P17 who stated:

I want to have the same gesture for next and previous regardless if I am viewing search results, contacts, or photos [P17]

Search results and contacts were arranged in a vertical list, whereas photos were arranged in a horizontal list. The gesture for “next” was common to both lists.

Finally, for gestures designed to navigate content (e.g., scrolling or panning) movement of the viewport can occur in discrete steps or can be based on the amount of force occurring during the gesture. While the agreement among participants was not as strong as other themes, there was a majority agreement that discrete navigation was preferred. As stated by P9:

If it was continuous then I think it would be pretty hard when to determine when to stop...and if I was walking down the street I would have to pay close attention [to] when to stop. [P9]

This observation runs counter to common surface gesture design on touch-screen smartphones, where on-screen flicks typically map gesture speed to different scrolling and panning distances.

Feedback

While the goal of our experiment was to eliminate any feedback in order to observe participants’ unedited gestures, participants often commented on the need for feedback:

I suppose what I would expect no matter what gesture I would use is some kind of feedback, probably some auditory feedback since I wouldn’t necessarily be looking at the phone...just alert me that is what it is doing and give me a chance to back out because I can imagine doing [the gesture] by mistake. [P9]

Participants were also careful when designing gestures to ensure that any visual feedback displayed on the screen would be visible during execution of the gesture.

The problem that any gesture that requires anything extended while you’re not looking at the screen...you are then losing feedback so it seems like it’s undesirable. [P10]

Since the tasks selected in our trial required the user to interact with content on the screen after completing the gesture—map navigation, navigating previous and next in lists, etc.—it was also important for participants to be able to view the screen while performing the gesture, especially

when part of the user experience is the interaction with the content. For example, P8 states:

...with photos usually there is this nice experience of like transitioning between one photo or the next, so I don't want to twitch because then I miss it. I want something that keeps the display facing me. [P8]

Motion Gesture Taxonomy

Given our understanding of the heuristics our participants applied to gesture design, the second question we explored is the set of parameters manipulated by our participants. We constructed a taxonomy for motion gestures using the 380 gestures collected that contains two different classes of taxonomy dimensions: *gesture mapping* and *physical characteristics*. *Gesture mapping* involves how users map motion gestures to device commands. These include the *nature*, *temporal* and *context* dimensions of the gesture. *Physical characteristics* involve characteristics of the gestures themselves: the *kinematic impulse*, *dimensionality*, and *complexity*. The full taxonomy is listed in Table 2.

Gesture Mapping

The *nature* dimension defines the mapping of the gesture to physical objects. One can view the gesture in a number of ways, specifically:

- *Metaphor*: The gesture is a metaphor of acting on a physical object other than a phone (a microphone, an old-fashioned phone).
- *Physical*: The gesture acts on the content/object itself (direct manipulation).
- *Symbolic*: The gesture visually depicts a symbol. For example, drawing the letter B with the device.
- *Abstract*: The gesture mapping is arbitrary.

The *temporal* dimension describes if the action on an object occurs during or after a gesture is performed. A gesture is categorized as *discrete* if the action on the object occurs after completing the gesture. Examples of discrete gestures include answering and making a call. During a *continuous* gesture, action occurs during the gesture and is completed upon the completing of the gesture. For example, map navigation tasks were typically considered continuous gestures by our participants.

The *context* dimension describes whether the gesture requires a specific context or is performed independent of context. For example, placing the phone to the head to answer a call is an *in-context* gesture, whereas a shaking gesture to return to the home screen is considered an *out-of-context* gesture.

Physical Characteristics

Since motion gestures are physical, it is appropriate to classify the gestures in reference to their *kinematic* properties. The kinematic impulse dimension segments gestures into three categories, represented by the range of jerk (rate of change of acceleration) applied to the phone

Taxonomy of Motion Gestures		
Gesture Mapping		
Nature	Metaphor of physical	Gesture is a metaphor of another physical object
	Physical	Gesture acts physically on object
	Symbolic	Gesture visually depicts a symbol
	Abstract	Gesture mapping is arbitrary
Context	In-context	Gesture requires specific context
	No-context	Gesture does not require specific context
Temporal	Discrete	Action occurs after completion of gesture
	Continuous	Action occurs during gesture
Physical Characteristics		
Kinematic Impulse	Low	Gestures where the range of jerk is below 3m/s^3
	Moderate	Gestures where the range of Jerk is between 3m/s^3 and 6m/s^3
	High	Gestures where the range of Jerk is above 6m/s^3
Dimension	Single-Axis	Motion occurs around a single axis
	Tri-Axis	Motion involves either translational or rotational motion, not both.
	Six-Axis	Motion occurs around both rotational and translational axes
Complexity	Simple	Gesture consist of a single gesture
	Compound	Gesture can be decomposed into simple gestures

Table 2. Taxonomy of motion gestures for mobile interaction based on collected gestures.

throughout the gesture. A low impulse gesture represents a gesture where the range of jerk over the gesture is below 3m/s^3 . A high impulse gesture is one where the range of jerk is relatively high, larger than 6m/s^3 over the gesture. An example of a high impulse gesture would be a forceful

shake. Gestures falling in between the range are classified as having a moderate kinematic impulse. The three categories and their respective cut-offs were determined by creating a histogram of the collected gestures by the rate of jerk and identifying clusters.

The *dimension* of a gesture is used to describe the number of axes involved in the movement. Many gestures, including flicks and flips of the phone involve single-axis motion. Others, for example zooming using a magnifying glass metaphor, require users to translate the phone in 3D space. Gestures that are either translations or rotations are tri-axis gestures. Still other gestures, for example ending a call by “hanging up” the phone, require users to both translate and rotate the device around its six degrees of freedom.

The *complexity* dimension relates to whether the proposed gesture is a *compound* gesture or a *simple* gesture. We define a compound gesture as any gesture that can be decomposed into simple gestures by segmenting around spatial discontinuities in the gesture. Discontinuities can include inflection points, pauses in motion, or corners.

Figure 1 illustrates the breakdown of the 380 gestures collected during the study using our taxonomy. As shown in the figure, gestures tended to be simple discrete gestures involving a single axis with low kinematic impulse.

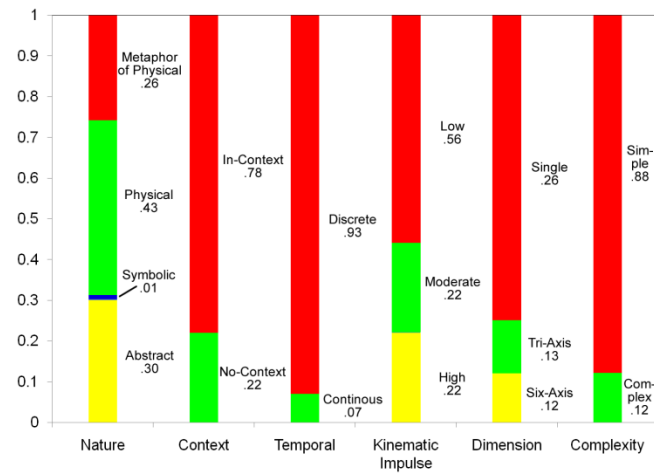


Figure 1. Percentage of gestures in each taxonomy category.

A User-defined Gesture Set

Using the gestures collected from our participants, we generated a user-defined gesture set for our specified tasks. For each task, identical gestures were grouped together. The group with the largest size was then chosen to be the representative gesture for the task for our user-defined gesture set. We call this gesture set both our *consensus set* and our *user-defined gesture set* interchangeably.

To evaluate the degree of consensus among our participants and compare our gesture set to Wobbrock et al. [28], we adopted the process of calculating an *agreement score* for each task [27,28]. An agreement score, A_t , reflects in a

single number the degree of consensus among participants. Wobbrock [27] provides a mathematical calculation for agreement, where:

$$A_t = \sum_{P_i} \left(\frac{P_i}{|P_t|} \right)^2$$

In Equation 1, t is a task in the set of all tasks T , P_t is the set of proposed gestures for t , and P_i is a subset of identical gestures from P_t . The range for A is $[0, 1]$.

As an example of an agreement score calculation, the task *answer the phone* had 4 groups with sizes of 17, 1, 1, and 1. Therefore, the agreement score for *answer the phone* is:

$$A_{\text{answer phone}} = \left(\frac{17}{20} \right)^2 + \left(\frac{1}{20} \right)^2 + \left(\frac{1}{20} \right)^2 + \left(\frac{1}{20} \right)^2 = 0.73$$

Figure 2, illustrates the agreement for the gesture set developed by our participants. Agreement scores from our user-defined motion gestures are similar to those shown for Wobbrock et al.’s gesture set for surface computing [28]. As shown by their agreement scores, there was not a consensus on a motion gesture for *switching to next application*, *switching to previous application*, and *act on selection* tasks. Therefore, we did not include gestures in the user-defined set for these tasks. The resulting user-defined set of motion gestures is shown in Figure 3.

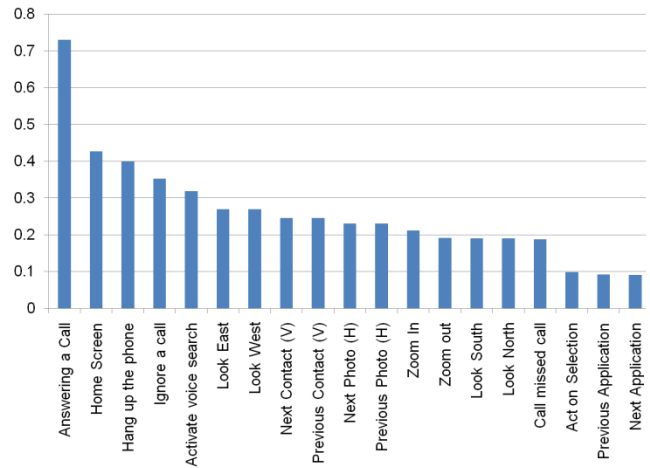


Figure 2. Agreement for each task sorted in descending order.

Subjective Ratings of the User-Defined Gesture Set

Recall that after designing a gesture for a particular task, participants rated the goodness of fit, ease of use, and how often the participant would use the gesture assuming it existed. Consider two sets of gestures. The first set is those gestures in our user-defined gesture set, i.e. those that were specified by a plurality of participants for each task. The second set includes all other gestures, i.e. those that are not part of our consensus set.

Comparing the subjective ratings, we find subjective ratings on *goodness of fit* to be more highly rated for our user-defined gesture set than for those gestures not in the consensus set ($X^2=12.85$, $p < 0.05$). However, we did not

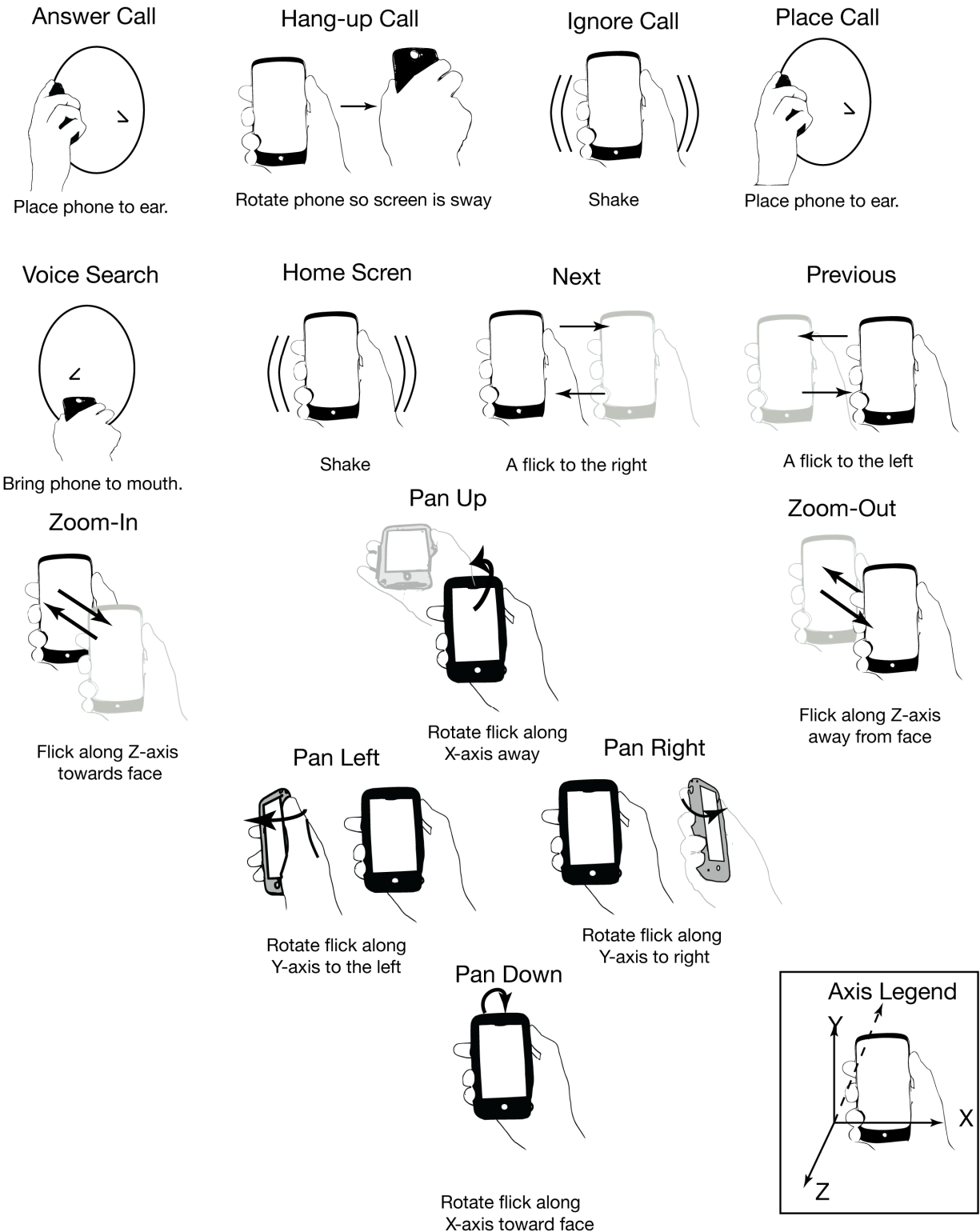


Figure 3. The user-defined motion gesture set. A *flick* is defined by a quick movement in a particular direction and returning to the starting position. Since the gestures for next and previous did not differ regardless how the task was presented (i.e. a vertical or horizontal list) we simplify our presentation by presenting the gestures under a single heading. The tasks of navigating to previous application, navigating to next application, and act on selection were not included in the gesture set due to the lack of agreement between participants.

find significant differences for *ease of use* or *frequency of use* between the two groups.

Task Coverage

Although great care was taken to create a list of tasks that would be representative of the tasks users perform on their device, it is possible that some potential tasks were not represented. To compensate for any neglected tasks, at the end of the interview we gave participants the opportunity to suggest tasks that would benefit from motion gesture interaction. While we did receive some suggestions, all suggestions were specific to an application, for example, a web browser. In addition, participants often commented that they would reuse previous gestures that were designed for the same purpose. For example, to mimic the back button in the web browser users suggested the same gesture as navigating to a previous photo or contact. Therefore, while we did not address all applications with our scenarios, we did address the majority of actions commonly used on mobile smartphones. As a result, our user-defined gesture set can be used to inform the design of motion gestures for a majority of the tasks within an application. The generalizability of many of the user-specified gestures allows for consistency across applications, which is important for learnability and memorability [29].

DISCUSSION

In this section we discuss the broader implications of our results for motion gesture design, mobile devices, and gesture interaction.

Supporting Motion Gesture Design

To support applications designers, motion gesture design software and toolkits should provide easy access to the gestures described in the user-defined set. Application designers may also wish to specify their own gestures, so design software and toolkits should also allow the creation of new motion gestures based on the heuristics presented above. Finally, while many tasks had good agreement scores for their user-specified gestures, some did not. For tasks with poor agreement scores, gesture toolkits should allow end-user customization.

Implications for System Design

The gestures in the user-defined gesture set and the themes that emerged from the study provide several challenges for designers of mobile phones. A major theme that emerged was that gestures should mimic normal use. In addition, as shown in Figure 1, a majority of the gestures collected during the study were classified as having a low kinematic impulse. The difficulty of using gestures with a low kinematic impulse and that mimic normal use is that these gestures are often difficult to distinguish from everyday motion. This can result in a high false positive rate and a high level of user frustration. As well, gestures with low kinematic impulse may be difficult to differentiate from one another using the current sensors in smartphones.

Despite the drawbacks associated with the observation that many user-specified motion gestures exhibit low kinematic

impulse and mimic normal use, there are benefits in understanding the natural characteristics of user-specified gestures. First, it may be possible to develop a delimiter, for example a physical button to push or an easy-to-distinguish motion gesture [20] to segment motion gestures from everyday device motion. Second, understanding the physical characteristics of end-user specified gestures gives system designers specific requirements to build towards. These can include adding additional sensors to infer context or using more sensitive sensors to distinguish between different low kinematic impulse motion gestures.

Implications for Gesture Interaction on Smartphones

During our study we asked participants how often they would use a motion gesture to accomplish a task. As we reported above, ratings did not differ depending on if the gesture designed by the participant was a member of the consensus group or not. In both cases, participants were very receptive to using motion gestures; only 4% of all responses indicated that participants would never use the motion gesture. In contrast, 82% of the responses indicated they would use the motion gesture at least occasionally. This result supports the notion that the use of motion gestures can substantially alter how users interact with their mobile phones. By providing motion gestures as an additional input modality, motion gestures can be used to simplify interaction (such as when answering the phone) or to enable interaction when users are unable to interact with the device using surface gestures (such as when wearing gloves).

Implication for the Design of 2D and 3D Gestural Interfaces

When examining the user-specified gesture set for surface computing developed by Wobbrock et al. [24], designers may feel that the existence of this gesture set was a singular event. In other words, something “special” about the intersection of surface computing with two-dimensional gestures permitted the creation of this gesture set. Our research indicates that this is not the case. For a different gestural paradigm, motion gestures, and for a different computing paradigm, mobile computing, another user-specified gesture set was created using a guessability study. As with surface gestures, our agreement scores vary from task to task. However, the extent of the between-participant agreement on gestures and mappings is still highly significant.

While gestures and mappings agree for surface and mobile computing paradigms, there are still open questions. Does agreement persist for other paradigms? For example, what about motion gesture interfaces where users control devices from afar using a device, an object, or their hands? What about using “scratches,” another form of surface gesture, to issue commands [7]. Our work suggests that conducting a guessability study with users before specifying gesture sets and mapping will significantly inform the design of gestures in these domains.

Social Acceptability of a User-Defined Gesture Set

Recent work by Rico and Brewster [19] and Montero et al. [14] explored the social acceptability of performing motion gestures in public places. The researchers found that a participant's rating of the social acceptability of a gesture was influenced by whether they believed a bystander could interpret the intention of the gesture. Given these findings, gestures in the consensus set (or gestures that mimic gestures in the set) should be more socially acceptable than gestures not in the set as a result from bystanders being able to interpret the meaning of the gesture. We plan on validating this hypothesis in future work.

FUTURE WORK

A limitation of our study is that our participants were educated adults who lived in a Western culture. It is quite possible that the gestures are influenced by the culture. For example, gestures such as previous and next are strongly influenced by reading order. In future work we would like to validate the user-defined gesture set with new participants from other user demographics and cultures.

We are also exploring tools to help developers select and evaluate gestures based on our taxonomy. In addition, we are exploring the possible use of online tools to allow the developer community to continue to revise and expand the user-defined gesture set as the tasks that users wish to accomplish on mobile devices change.

CONCLUSION

In this paper, we described the results of a guessability study for motion gestures. We show that for a subset of tasks that encompass actions with the device there is broad agreement on the motion gestures used to invoke these tasks. As a result of commonalities in gestures and their mappings, we present design heuristics and a taxonomy that inform motion gesture design for mobile interaction. Finally, we highlight the significant effect of this work on the paradigm of gestural interaction.

REFERENCES

1. *Android Open Source Project*. Google Inc.
2. *Etch A Sketch*. Ohio Art.
3. Ashbrook, D. and Starner, T. MAGIC: A Motion Gesture Design Tool. *Proceedings of CHI '10*, ACM (2010), 2159-2168.
4. Bartlett, J.F. Rock 'n' Scroll Is Here to Stay. *IEEE Comput. Graph. Appl.* 20, 3 (2000), 40-45.
5. Good, M.D., Whiteside, J.A., Wixon, D.R., and Jones, S.J. Building a user-derived interface. *Communications of the ACM* 27, 10 (1984), 1032-1043.
6. Harrison, B.L., Fishkin, K.P., Gujar, A., Mochon, C., and Want, R. Squeeze me, hold me, tilt me! An exploration of manipulative user interfaces. *Proceedings of CHI '98*, ACM Press/Addison-Wesley Publishing Co. (1998), 17-24.
7. Harrison, C. and Hudson, S.E. Scratch input. *Proceedings of UIST '08*, (2008), 205.
8. Hartmann, B., Abdulla, L., Mittal, M., and Klemmer, S.R. Authoring sensor-based interactions by demonstration with direct manipulation and pattern recognition. *Proceedings of CHI '07*, (2007), 145.
9. Hinckley, K., Pierce, J., Sinclair, M., and Horvitz, E. Sensing techniques for mobile interaction. *Proceedings of UIST '00*, ACM (2000), 91-100.
10. Hutchins, E., Hollan, J., and Norman, D. Direct Manipulation Interfaces. *Human-Computer Interact* 1, 4 (1985), 311-338.
11. Jones, E., Alexander, J., Andreou, A., Irani, P., and Subramanian, S. GesText: Accelerometer-based Gestural Text-Entry Systems. *Proceedings of CHI '10*, (2010).
12. Liu, J., Zhong, L., Wickramasuriya, J., and Vasudevan, V. User evaluation of lightweight user authentication with a single tri-axis accelerometer. *Proceedings of MobileHCI '09*, ACM (2009), 1-10.
13. Mignot, C., Valot, C., and Carbonell, N. An experimental study of future "natural" multimodal human-computer interaction. *Proceedings of INTERACT '93 and CHI '93*, (1993), 67-68.
14. Montero, C.S., Alexander, J., Marshall, M.T., and Subramanian, S. Would you do that? *Proceedings of MobileHCI '10*, (2010), 275.
15. Morris, M.R., Huang, A., Paepcke, A., and Winograd, T. Cooperative gestures. *Proceedings of CHI '06*, (2006), 1201.
16. Morris, M.R., Wobbrock, J.O., and Wilson, A.D. Understanding users' preferences for surface gestures. *Proceedings of GI 2010, CIPS* (2010), 261-268.
17. Partridge, K., Chatterjee, S., Sazawal, V., Borriello, G., and Want, R. TiltType: accelerometer-supported text entry for very small devices. *Proc. UIST '02*, ACM (2002), 201-204.
18. Rekimoto, J. Tilting operations for small screen interfaces. *Proceedings of UIST '96*, ACM (1996), 167-168.
19. Rico, J. and Brewster, S. Usable gestures for mobile interfaces: evaluating social acceptability. *Proceedings of CHI '10*, ACM (2010), 887-896.
20. Ruiz, J. and Li, Y. DoubleFlip: A Motion Gesture for Mobile Interaction. *Proceedings of CHI '11*, ACM (2011).
21. Schuler, D. *Participatory design: principles and practices*. L. Erlbaum Associates, Hillsdale N.J., 1993.
22. Small, D. and Ishii, H. Design of spatially aware graspable displays. *CHI '97 extended abstracts*, ACM (1997), 367-368.
23. Tang, J. Findings from observational studies of collaborative work. *International Journal of Man-Machine Studies* 34, 2 (1991), 143-160.
24. Volda, S., Podlaseck, M., Kjeldsen, R., and Pinhanez, C. A study on the manipulation of 2D objects in a projector/camera-based augmented reality environment. *Proceedings of CHI '05*, (2005), 611.
25. Weberg, L., Brange, T., and Hansson, W. A piece of butter on the PDA display. *CHI '01 extended abstracts*, ACM (2001), 435-436.
26. Wigdor, D. and Balakrishnan, R. TiltText: using tilt for text input to mobile phones. *Proc. UIST '03*, ACM (2003), 81-90.
27. Wobbrock, J.O., Aung, H.H., Rothrock, B., and Myers, B.A. Maximizing the guessability of symbolic input. *CHI '05 extended abstracts*, (2005), 1869.
28. Wobbrock, J.O., Morris, M.R., and Wilson, A.D. User-defined gestures for surface computing. *Proceedings of CHI '09*, (2009), 1083.
29. Wu, M., Chia Shen, Ryall, K., Forlines, C., and Balakrishnan, R. Gesture Registration, Relaxation, and Reuse for Multi-Point Direct-Touch Surfaces. *Proceedings of TABLETOP '06*, 185-192.
30. Yee, K. Peephole displays. *Proceedings of CHI '03*, ACM (2003), 1-8.