



HAL
open science

Exploring At-Your-Side Gestural Interaction for Ubiquitous Environments

Shaishav Siddhpuria, Keiko Katsuragawa, James R Wallace, Edward Lank

► **To cite this version:**

Shaishav Siddhpuria, Keiko Katsuragawa, James R Wallace, Edward Lank. Exploring At-Your-Side Gestural Interaction for Ubiquitous Environments. DIS 2017 - ACM Conference on Designing Interactive Systems, Jun 2017, Edinburgh, United Kingdom. pp.1111-1122, 10.1145/3064663.3064695 . hal-01654892

HAL Id: hal-01654892

<https://inria.hal.science/hal-01654892>

Submitted on 4 Dec 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Exploring At-Your-Side Gestural Interaction for Ubiquitous Environments

Shaishav Siddhpuria¹, Keiko Katsuragawa¹, James R. Wallace², Edward Lank^{1,3}

¹Cheriton School of Computer Science, University of Waterloo, Canada

²School of Public Health and Health Systems, University of Waterloo, Canada

³CRISAL, University of Lille, France

{spsiddhp, kkatsura, james.wallace, lank}@uwaterloo.ca

ABSTRACT

Free-space gestural systems are faced with two major issues: a lack of subtlety due to explicit mid-air arm movements, and the highly effortful nature of such interactions. With an ever-growing ubiquity of interactive devices, displays, and appliances with non-standard interfaces, lower-effort and more socially acceptable interaction paradigms are essential. To address these issues, we explore at-one's-side gestural input. Within this space, we present the results of two studies that investigate the use of side-gesture input for interaction. First, we investigate end-user preference through a gesture elicitation study, present a gesture set, and validate the need for dynamic, diverse, and variable-length gestures. We then explore the feasibility of designing such a gesture recognition system, dubbed *WatchTrace*, which supports alphanumeric gestures of up to length three with an average accuracy of up to 82%, providing a rich, dynamic, and feasible gestural vocabulary.

ACM Classification Keywords

H.5.m. Information Interfaces and Presentation (e.g. HCI): Miscellaneous.

Author Keywords

Smartwatch; Gestures; Ubiquitous Computing; Large Displays

INTRODUCTION

Free-space gestural input is an active area of research for ubiquitous display interaction (e.g., [1, 19, 20]) These techniques enable users to quickly and seamlessly interact with nearby displays, and have many practical advantages. To name only a few: proprioceptive interaction allows one to avoid focusing on the physical input device; hand and arm gestures can naturally combine intended target and command into a seamless whole; given gestures' role in interpersonal communication, there is some consensus on gesture meaning.

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. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

DIS 2017, June 10 - 14, 2017, Edinburgh, United Kingdom

© 2017 Copyright held by the owner/author(s). Publication rights licensed to ACM. ISBN 978-1-4503-4922-2/17/06...\$15.00

DOI: <http://dx.doi.org/10.1145/3064663.3064695>

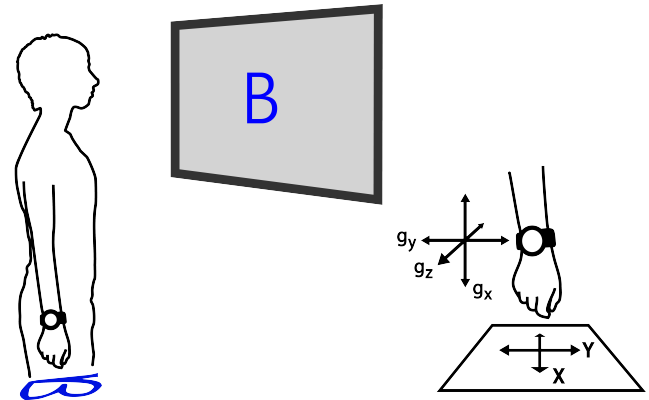


Figure 1. We conducted an elicitation study that focused on using at-your-side gestures to interact with ubiquitous resources, such as nearby displays. Data from the study informed the design of a proof-of-concept system, based on gravity sensor data collected from a smartwatch.

On the other hand, free space gestural input is not a panacea. It is an effortful form of communication, giving rise to terminology such as ‘gorilla arm syndrome’ and to metrics such as *consumed endurance* [11], which characterize the effort one must expend to support interaction. Alongside effort, free space communication is also noisy; it is difficult to exactly replicate motions in free space unless significant constraints are placed on the form of input path. Finally, sensing technology is challenging. Large scale gestures are easy to sense from afar, but small scale and subtle movements require spatial, set size, or kinematic constraints.

Recent work has identified two different aspects of interaction that might, conceivably address some of the challenges associated with free space gestural input. The first is a recent paper by Ruiz and Vogel on *Legacy Bias* [25]. In this work, Ruiz and Vogel explore natural free-space gestures to control an external display, specifically with the goal of addressing issues of fatigue. Using an elicitation study, they show that users can be primed to consider issues of fatigue in gesture elicitation by attaching simple weights onto their arms. The addition of these weights resulted in a dramatic decrease in consumed endurance measures for gestures elicited. The second piece of work that significantly influences us is work on the use of wearable devices to support gestural sensing. These

wearables take several forms – body mounted Leap Motion¹ devices [16], smartwatches [7, 13, 17, 31], special purpose input devices [10] – and permit gestural interaction with nearby computation via sensing technologies.

Overall, this research focuses on at-your-side gestural interaction (Figure 1), an interaction that minimizes *consumed endurance* by allowing users to interact with their arm at rest at their side. Alongside this, we focus on user consensus both in form and performance for these gestures. We explore whether there is a consensus set of gestures or, at least, a theme of consistency among end-users. Together, it is our hope that the work in his paper can add to research contributions such as the Gunslinger system [16], and the Legacy Bias observation [25] to further aid interaction designers in incorporating these highly subtle, low effort input gestures.

To satisfy these goals, this paper describes the results of two studies: first, an elicitation study explores gesture types for at-your-side interaction; then, a feasibility probe to explore the practicality of such a system. We provide a characterization of the types of gestures individuals create when gestures are elicited. One take-away from our elicitation study is the relatively low level of consensus in at-your-side interactions. Given the low-level of consensus, one question we had was whether or not it would be practical for at-your-side gestures to be user-specific. That is, could a system with low training overhead prove effective for an end user if they wished to define their own gesture set? The larger issue involved in this question is whether, with two or three templates, a user can create a sufficiently large gesture vocabulary and then reliably expect recognition from that vocabulary. This depends specifically on the relative precision (or lack of precision) in multiple instances of any gesture, such as how much neurophysiological and sensor noise is present in any captured signal and whether that noise is too high to support reliable recognition. Using off-the-shelf hardware and a simple template-based recognition algorithm, we demonstrate the feasibility of prototype systems that can accurately recognize complex gestures. We also develop a technique for multi-character gestures that allows users to leverage n -grams of gestures to create more complex commands.

RELATED WORK

Free-space interactions, such as hand gestures, have long been explored as a means of interaction with computers [28]. Such interactions have many natural benefits, such as ease of use, not requiring additional input hardware, and their universality. These benefits are particularly useful in ubiquitous and multi-device settings, and in supporting interactions as diverse as those with nearby appliances in the home, touch-free interaction with biomedical visualizations during surgery [22], or asking a nearby drone to take a ‘selfie’ [3].

We believe that a watch is an opportunistic form factor for sensing gestural input in these contexts. A watch has many advantages: as a wearable device it is ‘always on’ and ‘always available’, it is small, light, and conveniently located for easy access, and recent commercial products have made the form

factor inexpensive. In pursuit of leveraging these advantages, research has explored interaction paradigms [5, 13] and established toolkits for developing cross-device interfaces [12] for the growing ‘tsunami’ [8] of smart and embedded devices such as indoor lighting systems, thermostats and smoke detectors, and appliances.

Further, a smartwatch overcomes many of the limitations of existing gestural interaction methods [13]. For example, camera-based techniques suffer from occlusion and poor lighting conditions [10], and often require users to wear special markers as opposed to something they are already wearing. Further, these techniques require that body-tracking cameras be setup in any area where interaction may occur, and tracking performance is limited by the installed camera’s resolution [16]. Other sensor-based techniques often provide a fixed, relatively limited gestural vocabulary [27] where a gesture set is not readily modifiable or limited in personalization in a ubiquitous computing environment.

Recent work has explored the use of a smartwatch as a digital companion that supports gestural input. Notably, Porzi et al. [21] explore the use of a smartwatch in tandem with a smartphone’s camera, to assist visually impaired users avoid obstacles while walking. After developing an initial two-gesture prototype, follow-up work improved recognition to support up to 19 gestures [7]. Xu et al. [31] also leverage a smartwatch to support 37 finger and whole arm gestures. Where these projects have explored the use of a smartwatch to support visually impaired users, or interaction in specific contexts, we seek to make these devices more broadly useful, and to support interaction for all users and with any nearby device. Further tweaking off-the-shelf hardware to increase its sensory capabilities has shown to be promising in increasing the gestural vocabulary [14].

Minimizing Fatigue: Interaction at Your Side

Mid-air gesture interactions are prone to arm fatigue, a well-known phenomenon often referred to as the gorilla-arm effect [11]. The implications of arm fatigue are wide – ranging from general discomfort when interacting with systems, to systems going unused. For example, in a recent study of freehand interaction in the living room, 68% of gestures were changed over time, with researchers motivating these changes due to users experiencing fatigue during interaction [15]. Researchers have also explored the use of finger-based gestures for command invocation [4]. Such interaction techniques are potentially less fatiguing, but assume the use of supplementary finger and hand-tracking devices and cameras [16].

Because of the significance of fatigue in gestural input [25], a number of tools have been developed to help designers evaluate and minimize it. Notably, the Consumed Endurance (CE) metric provides a method to quantify arm fatigue via Kinect motion capture [11], and has been used to evaluate less fatiguing freehand gestures [25]. A key result of Hincapie-Ramos et al.’s evaluation is that interactions requiring lower shoulder torque will be less fatiguing. A natural corollary of the CE calculation is that, to prevent fatigue, gestures can take place at a user’s side instead of in the air in front of them, due to the negligible amount of shoulder torque.

¹www.leapmotion.com

GESTURE ELICITATION STUDY

Our research is motivated by the potential benefits of at-the-side interaction in terms of fatigue, and public performance/public exposure. We are not the first to advocate for this point-of-view: researchers have explored this input space for point-and-click interactions [16] to promote subtlety and lower fatigue. While point-and-click interactions are one mechanism for interacting with external computation, there may be limits to point-and-click in multi-device and command-driven environments. Furthermore it is not immediately clear that point-and-click interactions, even in environments where point-and-click might be possible, is the most natural form of interaction for at-the-side interactions.

One valuable tool for designers and researchers seeking to understand “naturalness”, “mental models”, or “consensus” in terms of input (or *cause*) vs invoked computational behavior (or *effect*) is the elicitation study. Elicitation studies have been commonly used in gestural input space to inform the design of gesture sets [24, 25, 29]. While the gesture sets generated by elicitation studies are not necessarily the definitive gesture set for use in an interaction paradigm, their value lies in adding to the understanding of the designer. The designer’s role is to resolve real world constraints, including data from the elicitation study, data on recognizer reliability, sensing capabilities, etc. to arrive at a gesture set that effectively supports input.

To drive our understanding of at-the-side gestural input, we performed an elicitation study to understand the types of gestures preferred by users for free space interaction in ubiquitous settings. We followed the protocol used in previous elicitation studies for free space gestures [24, 25, 29], where participants were invited into a laboratory and asked to indicate their preferred method of interacting with various nearby devices. The study provided an opportunity to elicit a gesture language appropriate for the subtle, at-your-side motions we wished to develop, and that provides a high degree of potential for personalization and expressivity.

We analyzed proposed gestures for general trends among participants: gesture length, type, complexity, whether participants chained or combined gestures into sequences, and recall rate for elicited gestures.

Design and Procedure

At the beginning of each experiment, after welcoming participants and obtaining consent, the researcher explained the study to the participant and asked them to wear a smartwatch and imagine interacting with a large display in a public setting.

We elicited gesture preference for 34 tasks (Table 1) using a think-aloud approach. For each task, participants were presented with an interface (Figure 2). They were told to select one gesture to perform the desired action and performed the gesture 4-5 times before finalizing their decision. Participants were instructed to choose any shape, symbol, movement or a combination of shapes, symbols, or movements for each task. The tasks were grouped by application and were presented one-by-one. We did provide some flexibility in ordering; for example, for tasks such as *Pan Up*, participants were free to clarify and choose a set of gestures for panning in the other

three directions: *Pan Down*, *Pan Left*, and *Pan Right* at the same time.

Images for each of the application interface were displayed on a 60" rear-projection system for simulating large display interaction (Figure 2). Although no sensor data was recorded, all the participants were required to wear a LG Watch R to simulate a working system.

Throughout the study, we did not want the participants to think of the capability of the gesture recognizers, and so we did not provide any feedback or confirmation to participants to remove the *gulf of execution* [18]. Instead, participants were instructed to only concern themselves with designing what they considered to be the most ‘natural’ interaction, assuming no limitations of such a gesture recognition system.

Once participants had selected a gesture, participants rated the gesture for its level of fatigue, whether it was a good match for the task, and whether they felt comfortable performing the gesture in public. These assessments were collected using a seven-point Likert-scale.

Finally, at the end of the study, a recall test was performed and participants were asked to comment on general design issues for free space interaction, such as “What constitutes a good gesture for a given task?”, “What are some other applications or use cases for a gesture-based input system?”, “Do you prefer alphanumeric gestures or ideographic gestures and why?”. Participants’ responses were later manually transcribed by the researcher. The video of the study was then transcribed to further identify common themes and trends using inductive, bottom-up reasoning. Each session took approximately one hour to complete.

Selection of Tasks

Our study uses a set of tasks (Table 1) previously used for gesture elicitation with mobile and whole-body interactions [10, 24]. We mimic the set of action metaphors for a new set of tasks in the context of familiar public displays: an airport departures screen, and a shopping mall kiosk. Our scenarios also included commonly used semi-public applications like Google Maps, Netflix, and Spotify. Recognizing that public display interfaces are inherently more dense with information in tabular or graphical layouts due to the real-estate available, we include an additional category of *Filter* tasks to encapsulate information selection in tabular, graphical, and list formats.

Participants

Twenty paid participants (10 female) between the ages of 21 and 55 ($\mu = 26.6$, $\sigma = 7.6$) were recruited. Participants were asked to use their dominant hand when eliciting gestures for each task, with three of the participants being left-handed. Every participant received \$10 remuneration. Participants were mainly university students in a Science, Technology, Engineering, or Mathematics field. We did not require or control for prior experience using or developing gestural systems.

Results

We now report results from the elicitation according to participant consensus, dictionary requirements, and recall and subjective measures.

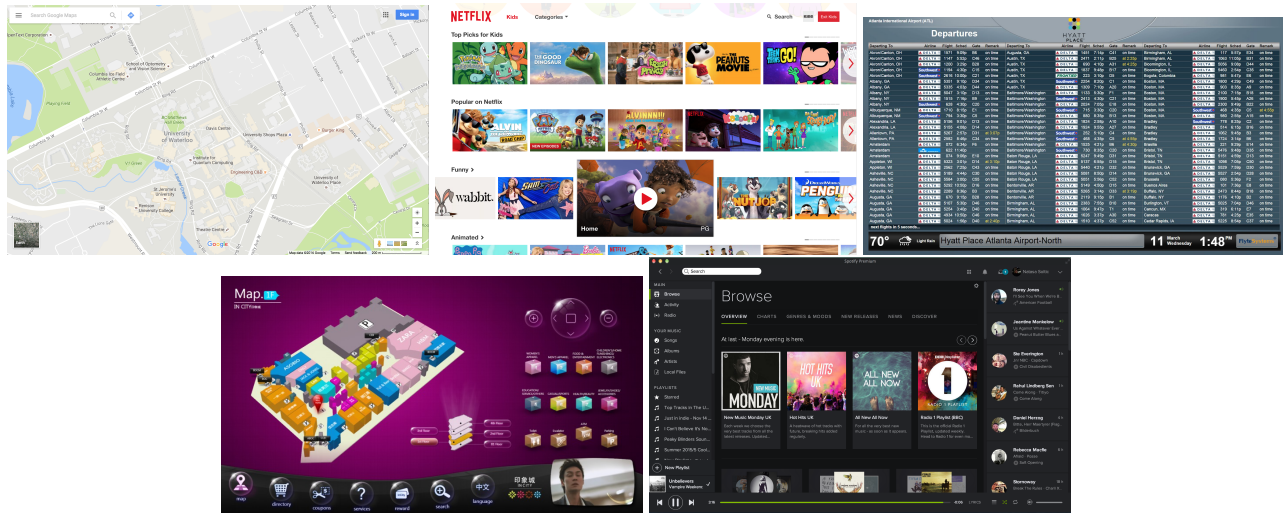


Figure 2. Interface mock-ups were displayed on a large projector for participants to imagine the task and elicit gestures. From left to right: Google Maps, Netflix, Airport Departures (hospitalitynet.org), Mall Display (gezhang.com), Spotify.

Context	Type		
	Navigation	Action	Filter
Google Maps	Zoom In/Out Pan Up/Down Pan Left/Right	Toggle View Show Menu Activate Search	Reset Location
Netflix	Select Next/Previous Page Left/Right	Play Item Activate Search Exit to Home Screen	Jump to Movie
Mall Display	Floor Up/Down		Jump to Store Jump to Floor Level
Airport Departures			Jump to Flight Filter Delayed Flights Clear Filter
Spotify	Next/Previous Song	Play Item Pause Item Replay Item Volume Up/Down Favorite Item	Jump to Menu Item

Table 1. The 34 tasks presented to participants during the elicitation study. Tasks were categorized by type: Navigation, Action, or Filter.

Analysis of Agreement

We first grouped together gestures that were visually identical — gestures described by participants with the same visual trace. For example, a ‘straight vertical line’ gesture trace would be identical when drawn either up to down or down to up. However, it would be different from a ‘horizontal line’ because the end trace of the two gestures visibly differ if drawn on a piece of paper. We then evaluated the degree of consensus among all the participants by using Wobbrock et al.’s methodology [29]. An agreement rate, AR is computed for each task as a single value between [0, 1] where a high value means that many participants chose the same gesture, and a low value means fewer participants chose the same gesture, hence more diversity in the gesture. Previous gesture elicitation studies [23, 25, 29] have used the same methodology to derive gesture sets across different domains.

Vatavu et al. [26] also provide guidance for interpreting AR, with suggested levels of low (≤ 0.1), medium (0.1 – 0.3), high (0.3 – 0.5), and very high (≥ 0.5) agreement. According to

this, we found a large amount of disagreement in the gestures, with more than 90% of the gestures having agreement values below 0.3, with a mean agreement value of 0.14 ($\sigma = 0.10$). The three tasks with high agreement were for *Play Song* (0.43), *Jump to Movie* (0.34), and *Jump to Floor Level* (0.30). AR for all tasks included in our study are shown in Figure 3.

Motion Gesture Taxonomy

In order to define the at-the-side gesture paradigm in the context of discount sensory capabilities of modern smartwatches, we fixed several taxonomy dimensions as compared to other, more free-form elicitation studies [23, 29, 25]. Specifically, the elicited gesture space was constrained to the two-axis plane parallel to the ground at one’s side, and the gesture context was fixed to the application view presented to the participant at that moment. Furthermore, the participant was not allowed to create a pose using their fingers as only the wrist movement pattern, forming a gesture trace was considered to be valid. Analyzing the elicited data, we defined the gesture taxonomy for at-the-side gesture paradigm with two dimensions: nature, and complexity.

In the nature dimension, we characterized the gestures as either ideographic (a shape or movement), alphanumeric (a stylized letter or number), or mixed (a combination of ideographic and alphanumeric). In order to better understand the diversity of gestures, we examined the gesture vocabulary of participants, i.e. how many unique gestures were used within the set of gestures elicited from a participant. We define unique gestures as follows: If the gestures elicited by the participants were to be drawn on a piece of paper, two gestures would be considered unique if they are either: a) visually different, b) oriented differently, or c) constructed in a different order of strokes. As an example, a list of compound gestures: “M”, “MV”, “VV”, and “MM” would only have two unique gesture symbols in the dictionary; “M”, and “V”. Alongside the gesture vocabulary size, we also examined the types of gesture used by participants. On average, the number of unique ideographic symbols per participant was 19.6 ($\sigma = 4.3$) across all 34 tasks.

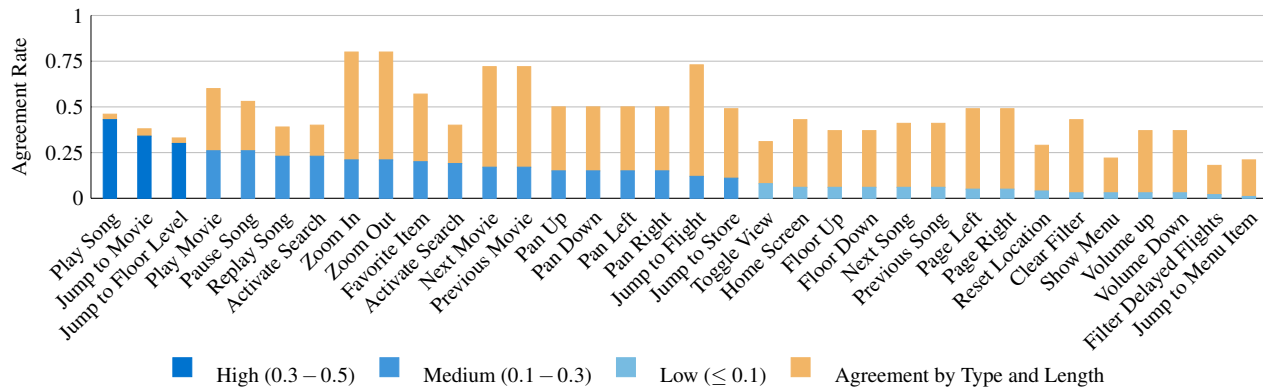


Figure 3. Agreement Rates (AR) for suggested gestures in terms of its visual trace, by each of the 34 tasks included in our elicitation study. Gestures ranked by agreement rating, with Vatavu and Wobbrock’s high, medium, and low agreement categories indicated by color. Orange bars show agreement scores when gestures are placed within more general groups of type and length.

For all of the *filter* tasks, the participants naturally proposed either completely alphanumeric or a mixture of ideographic-alphanumeric gesture sequences. For example, the elicited gestures for the *Jump to Movie* task in the Netflix grid-based layout proposed either writing out the letters of the movie title, or by specifying the row and column number corresponding to its position in the grid. Overall, 65.6% of collected gestures were classified as ideographic, 25.9% as alphanumeric, and 8.53% comprised a combination of the two.

For the *complexity* dimension, we categorized each gesture’s length as either “Single”, “Double” or “Triple+” to denote the conceptual length, as participants extended the vocabulary through gesture symbol concatenation and, at times, would create a gesture command using one symbol from the vocabulary. However, it was also quite common for participants to create two and three gesture sequences for gesture input. We use the term *conceptual length* to describe the number of atomic shapes used to create a gesture. For the purpose of analysis, we defined the concept of a shape as a series of curves and strokes being drawn in a specific order. The positioning of a stroke is also dependent on the preceding stroke. If either the position or the order of the strokes is independent of the preceding stroke, then it is marked as a natural delimiter among two atomic constructs. For example, if a participant designed a gesture that was “a house shape” comprised of a triangle on top of a rectangle, the gesture would have a conceptual length of 1. This is because of the co-dependent layout constraint of the triangle and rectangle. In contrast, a gesture that consisted of a triangle followed by a rectangle with no constraint on positioning would have conceptual length 2. As another example, a gesture sequence of “ZI”, contains two alphanumeric shapes that are only dependent on the order and not the location of the trace. Thus, this gesture has a conceptual length of 2. We define the conceptual length from the Overall, 71.8% of the gestures were of *single*, 20.5% were of *double*, and 7.65% were of triple or more conceptual length.

Using the *nature* and *complexity* taxonomy dimensions for a total of 9 groups (3 types \times 3 lengths), we calculated how much the participants agreed on a specific gesture type and length. Alexander et al. [1] and Chan et al. [4] have used

similar practices to reduce the diverse elicitation set to arrive at a consensus. By such loosening of the constraints on our classifications, we saw a mean agreement score of 0.46 across all the tasks ($\sigma = 0.16$) — a mean improvement of .3 across all tasks — with over 88% of the tasks having high to very high agreement values (Figure 3).

Based on this improved agreement, we choose a candidate gesture for each task in our elicitation study. Candidates were determined based on most frequent response within each of the Type \times Length categories. For example, for the *Zoom In* task, “Full circle clockwise” was the most frequent response within the most often suggested bin (Ideographic, Single gesture), and thus was chosen as our candidate gesture. Overall, this process yielded candidate gestures for each task, of which 76.5% were ideographic with a conceptual length of one. Results of this analysis are summarized in Table 2.

Recall and Subjective Measures

At the end of the study, we performed a quick test to see if the participants could recall the gestures they had chosen for each tasks. The participant’s ability to recall the gestures had no statistical significance with the gesture type or the length, the mean recall rate across all the tasks and all the participants was 90.8% ($\sigma = 6.23\%$). While further investigation is needed for a study of recall with respect to gesture sets of various lengths and types, this preliminary test hints towards participants having little difficulty with alphanumeric or ideographic gestures.

We gauged whether the participant was likely to use such a system in the wild by asking them to rate their own gestures on a 1-7 Likert scale in terms of naturalness, social fit, and fatigue with the value of 1 being the least amount of the respective characteristic. The mean of mean Likert rating for gesture naturalness was 5.32 ($\sigma = 0.41$), fatigue was 2.73 ($\sigma = 0.66$), and social awkwardness was 2.36 ($\sigma = 0.38$); participants designed gestures that were very natural, low fatigue, and generally socially acceptable. The most fatiguing gestures naturally correlated with the gesture length, *Jump to Store* had the most “Triple+” gestures and had the highest mean fatigue of 4.7.

Task	Preferred type	Preferred Length	Candidate Gesture(s)
Maps			
Zoom In	Ideographic	Single	Full circle clockwise
Zoom Out	Ideographic	Single	Full circle counterclockwise
Pan Up	Ideographic	Single	Line up
Pan Down	Ideographic	Single	Line down
Pan Left	Ideographic	Single	Line left
Pan Right	Ideographic	Single	Line right
Switch View	Ideographic	Single	Full circle (conflict)
Activate Search	Ideographic	Single	Magnifying glass (circle + line down)
Current Location	Ideographic	Single	Full circle (conflict)
Activate Left Menu	Alphanumeric	Single	Letter “M”
Netflix			
Select Item	Alphanumeric	Double	Row number and column number
Next Item	Ideographic	Single	Line right
Previous Item	Ideographic	Single	Line left
Page Left	Ideographic	Single	Chevron left
Page Right	Ideographic	Single	Chevron right
Activate Search	Ideographic	Single	Magnifying glass (circle + line down)
Play Item	Ideographic	Single	Triangle play symbol
Letter “H” or the letter “M”			
Airport Display			
Jump to Flight	Alphanumeric	Triple+	Airline code and flight number
Filter Delayed Flights	Mixed	Triple+	Magnifying glass and letters “DEL”
Clear All Filters	Ideographic	Single	Star symbol
Mall Display			
Change Floor Level	Alphanumeric	Single	Floor number
Go Up	Ideographic	Single	Chevron up
Go Down	Ideographic	Single	Chevron down
Jump to Store	Mixed	Triple+	Magnifying glass (circle + line down) and store letters
Spotify			
Play Item	Ideographic	Single	Triangle play symbol
Pause Item	Ideographic	Single	Two vertical parallel lines
Next Item	Ideographic	Single	Arrow left
Previous Item	Ideographic	Single	Arrow right
Replay Item	Ideographic	Single	Replay symbol (counterclockwise circle with an arrowhead)
Volume up	Ideographic	Single	Arrow up
Volume Down	Ideographic	Single	Arrow down
Jump to Menu Item	Mixed	Triple+	Left square bracket and item letters
Favorite Item	Ideographic	Single	Heart symbol

Table 2. Candidate gesture set for different applications as derived from agreement based on type and length, and the visual trace.

Preliminary analysis of recall and subjective measures highlights one key advantage of users designing their own gesture set: they can better tweak the gestures depending on their circumstances. While a gesture set designer has to optimize for the most common cases, users can optimize their gesture sets to their individual preference, a reason for highly diverse gestures and yet consistent recall.

Elicitation Study Take Aways

While the elicitation study provides a significant set of data points, the goal is to leverage that data to provide guidelines for the design of at-the-side gesture sets. An overview of the design implications from our elicitation study are as follows:

- Support a sufficiently large gesture set. Participants used, on average, approximately 20 unique gestures over the 34 tasks provided in this study. Given the restricted context of interacting with an external display, we believe that gesture sets in the 25 to 100 symbol set are minimally necessary.
- Consider user-specific versus designer-specified gesture sets. Overall, we see relatively low agreement scores across navigation, action, and filter task types. However, at the same time, we see high recall rates for users vis a vis their own gestures, despite specifying gestures for over 30 tasks. We believe that the combination of low inter-participant agreement and high recall rates is a key characteristic of domains that can benefit from allowing users to create their own gesture libraries.
- Include both “ideographic” and “alphanumeric” gestures. We found a combination of both ideographic or abstract gesture sets and alphanumeric characters as symbols in participants’ elicited gesture dialects. This is perhaps unsurprising: language is an important tool for expressing concepts, and, particularly in command-rich environments, text is often the best way to filter command lists. Consider, for example, invoking programs on modern operating systems, where short typed text strings filter command lists to a manageable length before users select from a list.
- Support multi-glyph gestures. One way that users naturally expand a command language is by combining symbols together into bigrams and trigrams; in our study, one-third of the hundreds of input gestures collected from participants were multi-symbol gestures. While support for bigrams and trigrams effectively expands the input language, it introduces challenges for symbol segmentation.
- Finally, we see a need for organic and dynamic input sequences. In public contexts, one may only infrequently, perhaps only once, invoke a specific command. As an example, locating a specific datum on a high density information display might require filtering and selecting, but that datum might only be relevant in the present interaction. Symbols that can be effectively combined into novel, single-use commands have high value in public contexts.

In the remainder of this paper, we present a proof-of-concept implementation of a side-gesture system, called *WatchTrace* which evaluates the potential of side-gesture input as an interaction paradigm. We acknowledge that the use of a prototype

system as a feasibility probe is not commonplace; often in Computer Science, the system is the contribution. Here we claim limited novelty in the system but novelty in the observations it allows us to make about the potential for this interaction modality. Alexander et al. [1] also take this approach, and describe its goals as “While [using naïve recognition to assess accuracy] would be impractical for real-world use, we were interested in determining if the gestures can be reliably classified, not whether our method is the technologically best implementation.”

Like [1], we take advantage of off-the-shelf recognizers simply to demonstrate that available tools can already address the lack of agreement found in our elicitation study. Noticing the general trends of ideographic and alphanumeric gestures and how they were combined, the feasibility study then covers a higher-level, language-driven approach that encompasses any concrete version of at-your-side gestures, which is a novel contribution that we maintain through both the pilot and feasibility study. Finally, we explore in detail our data to assess overall feasibility of the input space.

PILOT STUDY

To explore the feasibility of the smartwatch as a platform to detect motion gestures we conducted a pilot study. One requirement that we noted in our elicitation study was both ideographic and alphanumeric data were important to assess as input. We could, conceivably, have allowed users to define their own gesture sets, but that would tell us little about feasibility, as distinctiveness of gesture sets designed by different users would impact feasibility.

Selecting a Gesture Vocabulary

Inspection of our previously elicited gestures suggested that existing recognizers may already be well equipped for recognition. In particular, we noticed similarities between our elicited gestures and those previously studied in the \$1 [30] and Palm Graffiti sets [9]. For example, the consensus map navigation gestures (Table 2) closely match circles and arrows within the \$1 gesture set. Similarly, the “Chevron Up” and “Chevron Down” consensus gestures for the Mall display resemble \$1’s “v” gesture. Thus, we use the \$1 gesture set because it already covers many of the gestures suggested by our participants, many researchers have already assessed recognition accuracy against it, and it represents an ideographic gesture set of approximately the correct size.

Alphanumeric gestures were also among consensus gestures for all of our example applications other than Spotify; for these gestures, the Palm Graffiti set was used. We use Palm Graffiti because it is another instance of a unistroke gesture language, is well-established in practice, and its alphabetic correspondence means that if one wishes to extend the language, combining gestures into bigrams and trigrams has an intuitive mapping onto linguistic constructs.

Classification

We used the Dynamic Time Warping (DTW) algorithm, a simple template-based pattern matching technique widely used for speech recognition systems and other continuous data streams

[2]. Typically, DTW measures the similarity index between two temporal signals that can vary in time or speed and computes the warping distance to match the two signals by either stretching or compressing the time axis. The classifier then computes the relative similarity of the sample with all the templates in the model and makes a prediction based on the lowest DTW warp distance.

We applied matrix transformation and scaling on the gesture traces to normalize the size and position of the trace on the drawing plane. DTW was applied to the normalized sample traces and normalized template traces. Classification was performed on clean, segmented data in an off-line setting.

Experimental Task and Design

Participants from a local university performed an off-line recognition task. Participants were asked to wear a LG G Watch R on their dominant hand and instructed to keep their hand down, close to their body (e.g. Figure 1). Gestures to be performed were presented on a Nexus 5 smartphone’s display.

The smartwatch gave two vibration cues for each gesture: the first one is a signal for the participants to start performing the gesture; the second one marks the end of the gesture and participants were asked to finish their gesture before this second cue. The vibration cues were spaced 2 seconds apart. No feedback was provided to participants during the course of the study. Linear acceleration, gyroscope, and orientation and rotation vectors were logged on the smartphone at 200 Hz. In total, 16 gestures × 10 repetitions × 8 participants = 1,280 gestures were collected for the \$1 Gesture set and 26 × 10 × 8 = 2,080 gestures for Palm Graffiti.

Finally, we ran the gestures through our recognizer to assess accuracy readings for each gesture set. We did not optimize the recognizer; our goal was simply to test the potential of recognition and the need for training data.

Results

We evaluated recognizer accuracy using leave-*n*-out evaluation. Overall, the prototype achieved a recognition rate of 93.75% ($\sigma = 4.69$) for \$1 input and 89.23% ($\sigma = 4.78$) for Palm Graffiti input with maximum recognition occurring using 9 templates for each gesture (leave-one-out). The recognition rate for the \$1 Gesture set and Palm Graffiti with one template per gesture was 88.05% ($\sigma = 5.3$) and 82.3% ($\sigma = 6.2\%$) respectively (Figure 4). Recognition rate increases with the number of templates, but stabilizes with three templates.

Discussion

Our accuracy rate for a naïve implementation of DTW is promising. Our data indicate, first, that even with three templates, a user can expect to receive reasonably high accuracy with at-the-side gestures. Furthermore, in analyzing data from our recognizer, one of the attributes of the data that we noticed is that positional data, particularly as offset from gravity was most revealing in our input. Essentially, because the participant’s arm was hanging at rest at his or her side, we could view side-gesture input as a two-dimensional drawing on a plane, and we leverage this in the recognizer design in the next study we describe.

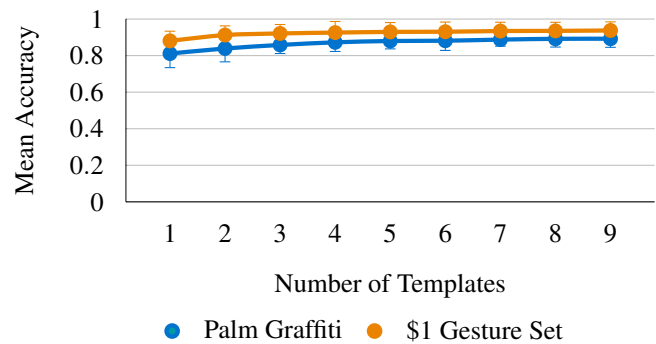


Figure 4. DTW recognition rates with a user-dependent template for the \$1 Gesture set and Palm Graffiti, by number of templates per gesture. Error bars show standard deviation.

Overall, while surface gesture error rates using prototyping recognizers are typically above 95% [30]), a recognition rate in excess of 90% for an unoptimized recognizer using only three training samples is, we feel, promising. There are many ways that gesture recognizer behavior can be optimized, including using *n*-best lists and confidence thresholds and reject decisions [6]. Our next study further validates the side-gesture interaction paradigm, exploring 3-gesture training set efficacy in more detail, incorporating a more ecologically valid assessment of input accuracy and combining this with an analysis of the feasibility of chaining gestures into bigrams and trigrams.

FEASIBILITY STUDY

To evaluate side gesture interaction in more ecologically valid settings, in an at-your-side paradigm, and to evaluate the potential of multi-glyph gesture recognition, we conducted a second laboratory study. As multi-glyph gestures, we used a set of 20 bigrams and trigrams as the vocabulary size from the list of most frequently appearing *n*-grams in the Wortschatz Leipzig Corpora Collection². Our system is designed specifically for contexts where gesture input is segmented from everyday movement of a device via a clutch [28] or delimiter [23]. In this follow-up study, we probe real-world challenges as our system does not assume cleanly segmented gesture data. Once data collection is started, we initially performed analysis on all input, including gesture and positional noise. During pilot testing we found that the embedded gravity sensor was sufficient for gesture recognition, and sensors such as a magnetometer are not embedded on every platform; thus only the gravity sensor was used in our proof-of-concept.

Apparatus and Data Collection

The experimental apparatus consisted of an LG G Watch R running Android Wear platform version 1.3, a Bluetooth-connected Nexus 5 smartphone with Android version 6.0.1, and a PC connected via TCP/IP. Motion input is captured with respect to gravity, allowing our system to work effectively using a 6-axis IMU. Gravity sensor data from the smartwatch was sampled at a rate of approximately 50Hz and forwarded to the PC via the smartphone, along with time-stamp information. On the PC, the sensor data was serialized into a JSON

²<http://corpora2.informatik.uni-leipzig.de/>

array and saved into a data file before being transferred to the recognizer application.

Data Mapping

We map the relative gravitational force onto a horizontal plane positioned at the user’s side and parallel to the ground. The relative movement of the user’s arm in this horizontal plane is represented in $[x_i, y_i]$ coordinates in the plane and is calculated as follows:

$$\begin{aligned} x_i &= x_{i-1} + (\Delta g_x) \\ y_i &= y_{i-1} + (\Delta g_y) \end{aligned}$$

Δg_y and Δg_z is calculated by taking the difference from the latest recorded sensor sample to the previously recorded sample, after being normalized. Note the transformation of coordinates on the smartwatch (g_y and g_z to the (x, y) plane). The z -component of the plane is not necessary for tracing the gesture because the z -vector with respect to the horizontal plane is aligned with the user’s arm, i.e. the two-dimensional plane’s z -axis and the smart watch’s x -axis are parallel (Figure 1). We perform our transformation to $[x, y]$ coordinates for convenience, to represent motion traces on an external display as in the accompanying video figure.

Generating the Template-based Model

During our experiment, we included a training phase to collect samples of the Palm Graffiti set [9] as 26 unigram gestures to seed the recognizer. The recognizer was trained by each participant so that the gestures being recognized were participant specific. For the purposes of the study, the converted gesture traces were not shown to the participants during both the training and the recognition phase as to not affect their natural behavior for performing gestures.

To create a set of gesture templates, we leveraged the DTW algorithm to find the best candidate input sample from every input gesture. For each training sample for a given gesture, we calculated its warp distance to all other training samples and then selected one single template to represent the gesture, the most representative of the candidate gestures, given by computing the minimum accumulative warp distance between every training sample collected for a single unigram.

To generate templates for the bigram and the trigram dictionary within our recognizer, we concatenated multiple unigram templates in order of the letter sequence, a common approach for template-based gesture recognizer systems. For example, to generate a template for the trigram “has”, we layered independent unigrams “h”, “a”, and “s” (Figure 5, right). We also interpolated data points between the end point of one unigram and the start point of the next, thus creating a single unistroke template that represented a multicharacter sequence, e.g. a bigram or trigram. The benefit of this approach is that once the recognizer is trained on the 26-gesture unigram alphabet, concatenation can yield an input language of up to 726 bigrams and of over 17,000 trigrams without additional training, being able to dynamically change its vocabulary without additional training data.

Gesture Recognition

Sensor data was recorded for each individual trial as a continuous stream including noise and non-relevant motion. We

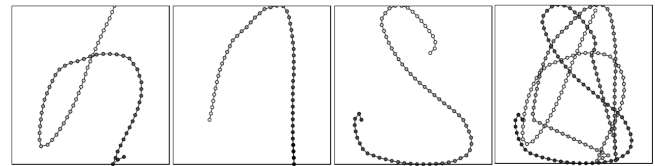


Figure 5. A template for a bigram or a trigram sequence is generated dynamically based on its individual unigram templates. The first three squares represent template for unigrams “H”, “A”, “S”. The last square represents the generated template for the trigram “HAS”.

perform segmentation of gestures (regardless of length) by detecting the idleness of the arm in the projected drawing plane. We use simple threshold-based segmentation to denote the beginning and end of a particular gesture. There are two components to the threshold, movement and time. We tuned our movement threshold from user input data from informal pilot studies. We use a threshold of 4 cm of arm movement in our horizontal plane, a displacement that seems to work well. From our initial pilot study, we found a temporal threshold of 1500ms to be a good threshold to distinguish short pauses between individual characters in a bigram or a trigram gesture and longer pauses between gestures. In summary, 1.5s of < 4 cm movement segments gestures. One single gesture could be a unigram character, a bigram character, or a trigram character.

Once the data was segmented and annotated as a gesture state, the recognizer performed DTW to the segmented data. Each sample was then annotated by the recognizer with the recognized gesture, trial instance information, and whether the prediction was successful or unsuccessful.

Participants

We recruited 12 participants (5 female) from a local university for a 30 minute session. All participants were right handed, and were instructed to wear the smartwatch on their non-dominant arm. Each participant received a \$5 honorarium.

Procedure

The procedure was as follows:

1. Participants were welcomed, provided background information on the purpose of the study, and signed an informed consent form. Next, participants were instructed on how to perform gestures using the smartwatch, asked to put the smartwatch on and adjust it for comfort. The participants were then asked to stand in front of a 60" rear-projection screen.
2. Participants then trained the recognizer. During the training phase, each participant was asked to repeat each of the single character Graffiti gestures (A to Z) 4 times, for a total of $26 \times 4 = 104$ trials. No feedback for the participant movement behavior was shown during the course of the study. To record clean sensor information for each motion gesture, participants delimited the start and end of each gesture, and were instructed to keep their arm idle when not performing a gesture. The collected data were used generate a template for the recognizer’s dictionary as described in the previous section. All samples provided as training data were, however, logged and preserved so that we could perform post-hoc analysis of variations in recognizer behavior.

3. Next, participants completed the recognition phase of the experiment. The recognition phase simulated a single gesture session where the user would perform gestures segmented by idle arm movement. The recognition phase comprised three stages of increasingly complex gestures: first unigram, then bigram, and trigram. We structured the study this way so that participants could familiarize themselves with the initial unigram gesture set before attempting bigram and trigram gestures. During each stage, the experimental software prompted participants to perform a random gesture of the specified length. If the gesture was recognized, the system would provide on-screen feedback by highlighting the gesture in green. If the gesture was not recognized, the user was notified by red-colored feedback and was given one additional attempt to perform the same gesture. If the gesture was not recognized in either attempt, the trial was marked as an error and the participant proceeded to the next trial. For this portion of the study, only the feedback for the gesture classification was shown to the participants, and not the actual gestural trace being recognized.

Results & Discussion

We performed a post-hoc analysis of collected gesture data to understand whether the recognition rates could be achieved without developing a special-purpose recognizer. We processed the collected data streams by resampling data-points to correct for issues such as jitter and repositioning movements. With minor data processing and the off-the-shelf DTW-based recognizer, mean recognition rates for the unigrams, bigrams, and trigrams were 82% ($\sigma = 5.71\%$), 80% ($\sigma = 3.63\%$), and 82% ($\sigma = 13.8\%$), respectively. We were also interested in seeing how “far off” the incorrectly recognized gestures from the template model in terms of the predicted gesture.

Taking the best two candidates from the recognizer classification, with the first candidate being the predicted, we saw an increase in accuracy of the unigrams to 92% ($\sigma = 3.92\%$), bigrams to 92% ($\sigma = 5.95\%$), and the trigrams to 90% ($\sigma = 10.15\%$). The results are summarized in Figure 6. This means that even when the gesture was wrongly classified, the right prediction was usually the second guess. In designing such a system for real-world use, we therefore suggest that strategies such as n -best lists can be taken advantage of to improve recognition rates. Additional techniques such as rejection sampling, confusion matrices, bi-level thresholding, and adaptive recognizers could be applied to further improve such systems in practice.

CONCLUSION

The goal of this paper was to explore at-your-side gestural interaction. The primary contributions of this work are contained in our elicitation study, which inferred low user agreement with respect to gesture set and informed the development of our at-your-side gesture prototype, *WatchTrace*. Overall, the elicitation study demonstrates the feasibility of the technique, particularly with respect to high user satisfaction, low fatigue, and high memorability for gestures created. It also demonstrates that there is relatively low agreement between users within this gestural input space and hints towards a large, adaptive vocabulary of gestures being necessary even within the scope of single-display interactions.

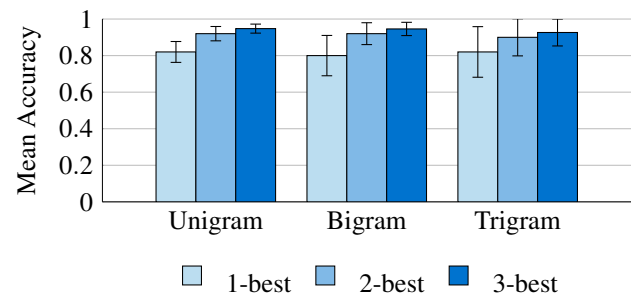


Figure 6. 1-best, 2-best, and 3-best recognition accuracy for 26 unigrams, 20 bigrams, and 20 trigrams with optimized classifier. Error bars show std deviation.

One challenge with the above list of high-level take-aways from our elicitation study is that, while these may be valuable guidelines for at-your-side gesture input, they provide little information about the feasibility of designing a system for side-gesture input in practice. To address this shortcoming, the second contribution of this paper is a feasibility probe: a prototype system that allowed us to assess whether or not side-gesture input recognizers can practically be constructed. Our pilot and feasibility studies demonstrate that one natural way to support this large gesture input space is via concatenation of gestures via bigrams and trigrams of simple gestures.

Our prototype system uses a commodity smartwatch to sense input, and it uses an off-the-shelf recognizer to analyze gestures. Overall, we use this feasibility probe to show that there is a high degree of consistency in user performance of at-your-side gestures. With four training samples, an effective recognition of over 90% is feasible, by incorporating standard error-thresholding techniques such as n -best lists, in a real-time recognition system that users train themselves, and generic template sets (collected from other users) also achieve high recognition accuracy. Alongside this, we demonstrate a technique for concatenation of simple gestures into bigram and trigram gestures and demonstrate high recognition accuracy for these more complex gestures.

At-the-side gestural input enjoys many significant advantages over other types of freehand gestures. It is low effort, as the gestures are performed with the arm at rest. Delimiters can be leveraged to discriminate input from everyday movement. It is easily sensed using commodity wearable devices, simplifying deployment. Its dialect is large thanks to the ability to leverage bi-grams and tri-grams. It can easily be combined with pointing through mode switches or by raising an arm to point at devices then lowering the arm to issue commands. Users rate the input as consistently low-effort and low-embarrassment for public use. Overall, the work described in this paper provides guidance to designers on how best to incorporate this gestural input modality as an interaction paradigm.

ACKNOWLEDGMENTS

We thank the Natural Sciences and Engineering Research Council of Canada (NSERC) and the Google Faculty Fellowship Award Program for funding this research.

REFERENCES

1. Jason Alexander, Teng Han, William Judd, Pourang Irani, and Sriram Subramanian. 2012. Putting Your Best Foot Forward: Investigating Real-world Mappings for Foot-based Gestures. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '12)*. ACM, New York, NY, USA, 1229–1238. DOI: <http://dx.doi.org/10.1145/2207676.2208575>
2. Donald J Berndt and James Clifford. 1994. Using Dynamic Time Warping to Find Patterns in Time Series.. In *KDD workshop*, Vol. 10. Seattle, WA, 359–370.
3. Jessica R. Cauchard, Jane L. E, Kevin Y. Zhai, and James A. Landay. 2015. Drone & Me: An Exploration into Natural Human-drone Interaction. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp '15)*. ACM, New York, NY, USA, 361–365. DOI: <http://dx.doi.org/10.1145/2750858.2805823>
4. Edwin Chan, Teddy Seyed, Wolfgang Stuerzlinger, Xing-Dong Yang, and Frank Maurer. 2016. User Elicitation on Single-hand Microgestures. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16)*. ACM, New York, NY, USA, 3403–3414. DOI: <http://dx.doi.org/10.1145/2858036.2858589>
5. Xiang 'Anthony' Chen, Tovi Grossman, Daniel J Wigdor, and George Fitzmaurice. 2014. Duet: Exploring Joint Interactions on a Smart Phone and a Smart Watch. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '14)*. ACM, New York, NY, USA, 159–168. DOI: <http://dx.doi.org/10.1145/2556288.2556955>
6. C Chow. 1970. On optimum recognition error and reject tradeoff. *IEEE Transactions on information theory* 16, 1 (1970), 41–46.
7. Gabriele Costante, Lorenzo Porzi, Oswald Lanz, Paolo Valigi, and Elisa Ricci. 2014. Personalizing a smartwatch-based gesture interface with transfer learning. In *Signal Processing Conference (EUSIPCO), 2013 Proceedings of the 22nd European*. IEEE, 2530–2534.
8. Alois Ferscha, Paul Lukowicz, and Franco Zambonelli. 2014. The Superorganism of Massive Collective Wearables. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication (UbiComp '14 Adjunct)*. ACM, New York, NY, USA, 1077–1084. DOI: <http://dx.doi.org/10.1145/2638728.2659396>
9. David Goldberg and Cate Richardson. 1993. Touch-typing with a Stylus. In *Proceedings of the INTERACT '93 and CHI '93 Conference on Human Factors in Computing Systems (CHI '93)*. ACM, New York, NY, USA, 80–87. DOI: <http://dx.doi.org/10.1145/169059.169093>
10. Faizan Haque, Matieu Nancel, and Daniel Vogel. 2015. Myopoint: Pointing and Clicking Using Forearm Mounted Electromyography and Inertial Motion Sensors. In *Proc. of CHI'15, the 33rd Conference on Human Factors in Computing Systems*.
11. Juan David Hincapié-Ramos, Xiang Guo, Paymahn Moghadasian, and Pourang Irani. 2014. Consumed Endurance: A Metric to Quantify Arm Fatigue of Mid-air Interactions. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '14)*. ACM, New York, NY, USA, 1063–1072. DOI: <http://dx.doi.org/10.1145/2556288.2557130>
12. Steven Houben and Nicolai Marquardt. 2015. WatchConnect: A Toolkit for Prototyping Smartwatch-Centric Cross-Device Applications. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI '15)*. ACM, New York, NY, USA, 1247–1256. DOI: <http://dx.doi.org/10.1145/2702123.2702215>
13. Keiko Katsuragawa, Krzysztof Pietroszek, James R. Wallace, and Edward Lank. 2016. Watchpoint: Freehand Pointing with a Smartwatch in a Ubiquitous Display Environment. In *Proceedings of the International Working Conference on Advanced Visual Interfaces (AVI '16)*. ACM, New York, NY, USA, 128–135. DOI: <http://dx.doi.org/10.1145/2909132.2909263>
14. Gierad Laput, Robert Xiao, and Chris Harrison. 2016. ViBand - High-Fidelity Bio-Acoustic Sensing Using Commodity Smartwatch Accelerometers. *UIST (2016)*, 321–333.
15. Sang-Su Lee, Jeonghun Chae, Hyunjeong Kim, Youn-kyung Lim, and Kun-pyo Lee. 2013. Towards More Natural Digital Content Manipulation via User Freehand Gestural Interaction in a Living Room. In *Proceedings of the 2013 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp '13)*. ACM, New York, NY, USA, 617–626. DOI: <http://dx.doi.org/10.1145/2493432.2493480>
16. Mingyu Liu, Mathieu Nancel, and Daniel Vogel. 2015. Gunslinger: Subtle Arms-down Mid-air Interaction. In *Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology (UIST '15)*. ACM, New York, NY, USA, 63–71. DOI: <http://dx.doi.org/10.1145/2807442.2807489>
17. Danial Moazen, Seyed A Sajjadi, and Ani Nahapetian. 2016. AirDraw: Leveraging smart watch motion sensors for mobile human computer interactions. In *2016 13th IEEE Annual Consumer Communications & Networking Conference (CCNC)*. IEEE, 442–446.
18. Donald A Norman. 1986. Cognitive engineering. *User centered system design: New perspectives on human-computer interaction* 3161 (1986).
19. Krzysztof Pietroszek, Anastasia Kuzminykh, James R. Wallace, and Edward Lank. 2014. Smartcasting: A Discount 3D Interaction Technique for Public Displays. In *Proceedings of the 26th Australian Computer-Human Interaction Conference on Designing Futures: The*

- Future of Design (OzCHI '14)*. ACM, New York, NY, USA, 119–128. DOI : <http://dx.doi.org/10.1145/2686612.2686629>
20. Krzysztof Pietroszek, James R. Wallace, and Edward Lank. 2015. Tiltcasting: 3D Interaction on Large Displays Using a Mobile Device. In *Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology (UIST '15)*. ACM, New York, NY, USA, 57–62. DOI : <http://dx.doi.org/10.1145/2807442.2807471>
 21. Lorenzo Porzi, Stefano Messelodi, Carla Mara Modena, and Elisa Ricci. 2013. A Smart Watch-based Gesture Recognition System for Assisting People with Visual Impairments. In *Proceedings of the 3rd ACM International Workshop on Interactive Multimedia on Mobile & Portable Devices (IMMPD '13)*. ACM, New York, NY, USA, 19–24. DOI : <http://dx.doi.org/10.1145/2505483.2505487>
 22. F. Ritter, C. Hansen, V. Dicken, O. Konrad, B. Preim, and H. o. Peitgen. 2006. Real-Time Illustration of Vascular Structures. *IEEE Transactions on Visualization and Computer Graphics* 12, 5 (Sept 2006), 877–884. DOI : <http://dx.doi.org/10.1109/TVCG.2006.172>
 23. Jaime Ruiz and Yang Li. 2011. DoubleFlip: A Motion Gesture Delimiter for Mobile Interaction. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '11)*. ACM, New York, NY, USA, 2717–2720. DOI : <http://dx.doi.org/10.1145/1978942.1979341>
 24. Jaime Ruiz, Yang Li, and Edward Lank. 2011. User-defined motion gestures for mobile interaction. In *Proceedings of the 2011 annual conference on Human factors in computing systems (CHI '11)*. ACM, New York, NY, USA, 197–206. DOI : <http://dx.doi.org/10.1145/1978942.1978971>
 25. Jaime Ruiz and Daniel Vogel. 2015. Soft-Constraints to Reduce Legacy and Performance Bias to Elicit Whole-body Gestures with Low Arm Fatigue. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI '15)*. ACM, New York, NY, USA, 3347–3350. DOI : <http://dx.doi.org/10.1145/2702123.2702583>
 26. Radu-Daniel Vatavu and Jacob O. Wobbrock. 2015. Formalizing Agreement Analysis for Elicitation Studies: New Measures, Significance Test, and Toolkit. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI '15)*. ACM, New York, NY, USA, 1325–1334. DOI : <http://dx.doi.org/10.1145/2702123.2702223>
 27. Hongyi Wen, Julian Ramos Rojas, and Anind K Dey. 2016. Serendipity: Finger Gesture Recognition using an Off-the-Shelf Smartwatch. (2016), 3847–3851.
 28. Daniel Wigdor and Dennis Wixon. 2011. *Brave NUI world: designing natural user interfaces for touch and gesture*. Elsevier.
 29. Jacob O. Wobbrock, Meredith Ringel Morris, and Andrew D. Wilson. 2009. User-defined Gestures for Surface Computing. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '09)*. ACM, New York, NY, USA, 1083–1092. DOI : <http://dx.doi.org/10.1145/1518701.1518866>
 30. Jacob O Wobbrock, Andrew D Wilson, and Yang Li. 2007. Gestures Without Libraries, Toolkits or Training: A \$1 Recognizer for User Interface Prototypes. In *Proceedings of the 20th Annual ACM Symposium on User Interface Software and Technology (UIST '07)*. ACM, New York, NY, USA, 159–168. DOI : <http://dx.doi.org/10.1145/1294211.1294238>
 31. Chao Xu, Parth H. Pathak, and Prasant Mohapatra. 2015. Finger-writing with Smartwatch: A Case for Finger and Hand Gesture Recognition Using Smartwatch. In *Proceedings of the 16th International Workshop on Mobile Computing Systems and Applications (HotMobile '15)*. ACM, New York, NY, USA, 9–14. DOI : <http://dx.doi.org/10.1145/2699343.2699350>