PalmSpace: Leveraging the Palm for Touchless Interaction on Public Touch Screen Devices

by

Pinku Deb Nath

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE

in

THE COLLEGE OF GRADUATE STUDIES

(Computer Science)

THE UNIVERSITY OF BRITISH COLUMBIA

(Okanagan)

April 2022

© Pinku Deb Nath, 2022
The following individuals certify that they have read, and recommend to the College of Graduate Studies for acceptance, a thesis/dissertation entitled:

**MAM** SPACE: **LEVERAGING THE PALM FOR TOUCHLESS INTERACTION ON PUBLIC TOUCH SCREEN DEVICES**

submitted by Pinku Deb Nath in partial fulfilment of the requirements of the degree of Master of Science.

Dr. Mohammad Khalad Hasan, CMPS, Irving K. Barber Faculty of Science  
**Supervisor**

Dr. Patricia Lasserre, CMPS, Irving K. Barber Faculty of Science  
**Supervisory Committee Member**

Dr. Ramon Lawrence, CMPS, Irving K. Barber Faculty of Science  
**Supervisory Committee Member**

Dr. Jahangir Hossain, School of Engineering, Faculty of Applied Science  
**University Examiner**
Abstract

The touchscreen is the primary solution to interact with public devices such as Automated Teller Machines (ATMs). However, the touch modality raises health concerns since users have to touch the screens, and therefore risk the spread of contagious diseases.

I designed PalmSpace, an alternate input technique leveraging users hand palms to interact with public devices. With PalmSpace, UI elements are mapped onto the users palms and can be accessed by touching various locations directly on the palm. I conducted a series of user studies to evaluate several design options, such as interface layout, item size, preferred item location, and suitable feedback for items. Based on the results, I designed PalmSpace and compared its performance with mid-air input. I showed that PalmSpace is a potential solution to interact with public devices without using their touchscreen. I concluded with design guidelines for using the palm as an alternative input space for touchscreen devices.
Lay Summary

Most public services such as ATMs can only be used with touchscreen or physical buttons. However, touchscreen interaction imposes health risks such as the spread of diseases. I investigated an alternative way of interaction without touching anything while being as natural as using a touchscreen. In my thesis, I propose PalmSpace, which uses the palm as an alternative input space to touchscreens. I ran several user studies to determine the best design choices for PalmSpace and measured performance based on the time taken to complete selection tasks. I integrated all the best choices and compared the performance of PalmSpace with the midair technique for purchasing train tickets. Results showed that purchasing tickets using PalmSpace is faster than interacting in midair. I further deployed the PalmSpace interface in public spaces and received positive feedback. I concluded with guidelines for designing interfaces such as PalmSpace in the future.
Preface

This thesis is the original, unpublished work by the author, Pinku Deb Nath. This study was granted approval by the University of British Columbia Behavioural Research Ethics Board- Certificate Number: H20-02018.
Table of Contents

Abstract ................................................................. iii
Lay Summary ............................................................. iv
Preface ................................................................. v
Table of Contents ....................................................... vi
List of Figures ........................................................... ix
Acknowledgements ...................................................... xiv
Dedication ............................................................. xv

Chapter 1: Introduction ................................................. 1

Chapter 2: Related Work ............................................... 6
  2.1 Mid-Air Interaction ............................................. 6
    2.1.1 Input Techniques ......................................... 6
    2.1.2 Feedback Methods ....................................... 9
    2.1.3 Applications ............................................. 12
    2.1.4 Acceptance ............................................... 13
  2.2 Body-Based Interaction ....................................... 14
    2.2.1 Input Techniques ......................................... 14
    2.2.2 Feedback Methods ....................................... 16
    2.2.3 Applications ............................................. 17
    2.2.4 Acceptance ............................................... 18

Chapter 3: Survey ....................................................... 19
  3.1 Participants ................................................... 19
  3.2 Procedure ..................................................... 19
# TABLE OF CONTENTS

3.2.1 Palm ........................................... 20  
3.2.2 Results ....................................... 22  
3.3 Discussion ...................................... 25  

Chapter 4: Design Exploration ............................... 26  
4.1 Pilot Study: Item Placement .......................... 27  
4.1.1 Participants .................................... 28  
4.1.2 Procedure ...................................... 28  
4.1.3 Results ......................................... 31  
4.1.4 Discussion ...................................... 31  

Chapter 5: Study 1 ........................................ 33  
5.1 Remediation Technique ............................... 33  
5.2 Grid Size .......................................... 34  
5.3 Grid Location ...................................... 35  
5.4 Participants ....................................... 35  
5.5 Apparatus ......................................... 35  
5.5.1 System Design .................................. 36  
5.6 Procedure ......................................... 39  
5.7 Results ........................................... 40  
5.8 Discussion ......................................... 42  

Chapter 6: Study 2 ........................................ 43  
6.1 Design Factors ..................................... 43  
6.1.1 Layout ......................................... 43  
6.1.2 Scanning Direction .............................. 43  
6.1.3 Integration ..................................... 45  
6.2 Participants and Apparatus ......................... 47  
6.3 Procedure ......................................... 47  
6.4 Design ............................................ 47  
6.4.1 Results ........................................ 48  
6.4.2 Target visit time ............................... 48  
6.4.3 Error rate ...................................... 49  
6.5 Preferences ........................................ 49  
6.6 Summary .......................................... 49  

Chapter 7: Study 3 ........................................ 50  
7.1 Participants and Apparatus ......................... 50  
7.2 Procedure ......................................... 51  
7.3 Design ............................................ 53
# TABLE OF CONTENTS

7.4 Results ................................................. 55
  7.4.1 Trial Time ........................................... 55
  7.4.2 Error Rate ......................................... 55
  7.4.3 Cursor Movement ..................................... 55
  7.4.4 Preference .......................................... 55
7.5 Summary ................................................. 55

Chapter 8: In The Wild Demonstration ................................. 57
  8.1 Conditions ............................................. 57
  8.2 Participants and Apparatus ................................ 58
  8.3 Procedure .............................................. 59
  8.4 Feedback and Observations ................................ 59
  8.5 Discussion .............................................. 60

Chapter 9: Design Guidelines, Limitations and Future Work ....... 62
  9.1 Design Guidelines ......................................... 62
    9.1.1 Palm as Input Space .................................. 62
    9.1.2 Location, Grid Size, and Layout ....................... 62
    9.1.3 Item Placement ....................................... 63
    9.1.4 Presentation and Remediation ............................ 63
    9.1.5 Discoverability ...................................... 63
  9.2 Limitations ............................................. 63
    9.2.1 Prototype .......................................... 63
    9.2.2 Number of Participants ................................ 64
  9.3 Future Work .............................................. 64
    9.3.1 Guidance to novice users ............................... 64
    9.3.2 Reducing Fatigue ..................................... 65
    9.3.3 Expanding interaction space ............................. 65

Chapter 10: Conclusion ........................................... 66

Bibliography .................................................. 68

Appendices .................................................... 87
  Appendix A: Design Exploration .................................. 87
    A.1 Combinations of on-screen items ............................ 87
    A.2 Demonstration .......................................... 87
      A.2.1 One group .......................................... 87
      A.2.2 Two groups .......................................... 87
## List of Figures

| Figure 1.1 | Input Spaces: (a) Palm, (b) On-Body, (c) Mid Air | 2 |
| Figure 1.2 | PalmSpace: an input technique leveraging users palms to interact with public devices | 3 |
| Figure 1.3 | Summary of design studies | 5 |
| Figure 1.4 | Summary of evaluation studies | 5 |
| Figure 3.1 | Dynamicity to initialize a window | 21 |
| Figure 3.2 | Remediation. (a) Hand-to-Screen, (b) Screen-to-Hand and (c) Vignette (combination of Hand-to-Screen and Screen-to-Hand). © Pinku Deb Nath, 2022 | 21 |
| Figure 3.3 | Relative Anchor using Screen-to-Hand remediation. Interface moves as the left hand moves from (a) left to (b) center and to the (c) right side. © Pinku Deb Nath, 2022 | 22 |
| Figure 3.4 | Triggers with the hand that is the anchor for Palm input space: (a) Thumb gesture, (b) Open Hand | 23 |
| Figure 3.5 | Triggers with the hand that is not the anchor for Palm input space: (a) Depth-tap, (b) Shooting gesture, (c) Pinch, and (d) Dwell | 23 |
| Figure 3.6 | Median user preferences for (a) number of hands and (b) gestures for initiating input space, (c) gestures for triggering selections and (d) overall preferences for input space. The lower values signify higher preference while the higher values signify lower preference | 24 |
LIST OF FIGURES

Figure 4.1 Combinations of on-screen items arranged in one group: (a) 2 items, horizontal layout, (b) 2 items, vertical layout, (c) 4 items, horizontal layout, (d) 4 items, vertical layout, (e) 4 items, grid layout with 2 rows, (f) 6 items, horizontal layout, (g) 6 items, vertical layout, (h) 6 items, grid layout with 2 rows, (i) 6 items, grid layout with 2 columns. 26

Figure 4.2 Combinations of on-screen items arranged in two groups with 2 items per group: (a) 2 items with horizontal layout per group and the groups arranged horizontally, (b) 2 items with horizontal layout per group and the groups arranged vertically, (c) 2 items with vertical layout per group and the groups arranged horizontally, (d) 2 items with vertical layout per group and the groups arranged vertically. 27

Figure 4.3 One group, horizontal (a, b, c) and vertical (d, e, f) arrangements of 6 items, (a, d) are initial states, (b, e) shows the result after a participant maps the items onto the palm image while (c, f) are the aggregated results of all the six participants’ input for the considered combinations. 28

Figure 4.4 Two groups representing horizontal group arrangement, 4 items arranged vertically within the two groups which are in horizontal orientation, (a) is the initial state, (b) shows the result after a participant maps the items onto the palm image while (c) are the aggregated results of all the six participants’ input for the considered combinations. 29

Figure 5.1 Techniques: (a) S2H Palm (b) H2S Palm (c) S2H Finger (d) H2S Finger. © Pinku Deb Nath, 2022. 34

Figure 5.2 System design of PalmSpace. 36

Figure 5.3 21 indices of hand landmarks starting from 0. 38

Figure 5.4 Mean target visit time for Palm and Finger across different Grid Size. Error bars: ± 2 S.E. 41

Figure 5.5 Target visit time in (a) 2×2, (b) 3×3 and (c) 4×4 grid cells. 41
Figure 6.1 Flow layout concept with respect to scanning direction. (a, b, c) Flow layout with left to right scanning direction, (a) Original arrangement of targets (A-E) in on-screen UI mapped onto the palm, (b) Rearrangement of the targets according to Flow layout with C, D, E in the last column and (c) Rearrangement of the targets according to Grid layout with B, C, D, E having the same dimensions and D starting from the second row. (d, e, f) Flow layout concept with right to left scanning direction, (d) Original arrangement of targets (A-E) in on-screen UI mapped onto the palm, (e) Rearrangement of the targets according to Flow layout with C, D, E in the last row and (f) Rearrangement of the targets according to Grid layout with A, B, D, E having the same dimensions, A and B starting from the second row with C occupying the whole first column.

Figure 6.2 Full integration (a) Left-to-Right with Flow Layout and (b) Right-to-Left with Grid Layout. © Pinku Deb Nath, 2022.

Figure 6.3 Partial integration: (a) Left-to-Right with Flow, (b) Right-to-Left with Grid. © Pinku Deb Nath, 2022.

Figure 6.4 (a) Mean target visit time and (b) mean error rates. Error bars: ± 2 S.E. (c) Median user ratings.

Figure 7.1 PalmSpace UIs: (a) Main, (b) Ticket Type, (c) Payment Amount, and (d) Payment Method.

Figure 7.2 UIs for Mid-Air Input Space, (a) Main, (b) Ticket Type, (c) Payment Amount, (d) Payment Method.

Figure 7.3 Mean (a) trial time and (b) cursor movement across two techniques. Error bars: ± 2 S.E. for Mid-Air and PalmSpace.

Figure 7.4 (a) Mean error rate across two techniques with error bars: ± 2 S.E. (b) Median preference rating for Mid-Air and PalmSpace.

Figure 8.1 In the wild demonstration in a (a) university building’s main hall, (b) in a restaurant, and (c) a user interacting with PalmSpace.
LIST OF FIGURES

Figure 8.2  Menu options: (a) online map, (b) weather report rendered in real-time inside iframe HTML element.
© Pinku Deb Nath, 2022. . . . . . . . . . . . . . . . . . . . . . . . . . 59

Figure A.1  Combinations of on-screen items arranged in two groups with 4 items per group: (a) 4 items with horizontal layout per group and the groups arranged horizontally, (b) 4 items with horizontal layout per group and the groups arranged vertically, (c) 4 items with vertical layout per group and the groups arranged horizontally, (d) 4 items with vertical layout per group and the groups arranged vertically, (e) 4 items with a grid layout of 2 rows per group and the groups arranged horizontally, (f) 4 items with a grid layout of 2 rows per group and the groups arranged vertically. 88

Figure A.2  Combinations of on-screen items arranged in 2 groups with 6 items per group: (a) 6 items with horizontal layout per group and the groups arranged horizontally, (b) 6 items with horizontal layout per group and the groups arranged vertically, (c) 6 items with vertical layout per group and the groups arranged horizontally, (d) 6 items with vertical layout per group and the groups arranged vertically, (e) 6 items with a grid layout of 2 rows per group and the groups arranged horizontally, (f) 6 items with a grid layout of 2 rows per group and the groups arranged vertically, (g) 6 items with a grid layout of 3 rows per group and the groups arranged horizontally, (h) 6 items with a grid layout of 3 rows per group and the groups arranged vertically. 89

Figure A.3  One Group, Grid Arrangement (a, b, c) and (d, e, f) represents two ways in which grid arrangements are considered for one group, (a, d) are initial states, (b, e) shows the result after a participant maps the items onto the palm image while (c, f) are the aggregated results of all the six participants’ input for the considered combinations. 90
Figure A.4  Two groups representing vertical group arrangement, 6 items arranged in a grid of 2 rows within the two groups, (a) are initial states, (b) shows the result after a participant maps the items onto the palm image while (c) are the aggregated results of all the six participants’ input for the considered combinations. . . . 91
Acknowledgements

I would like to express my gratitude to my supervisor Dr. Mohammad Khalad Hasan who provided me with his invaluable guidance, encouragement and generous help. I will always remember how patiently and calmly he would listen and acknowledge my ideas during our long discussions and brainstorming sessions at each phase of the thesis. His work ethics, immense knowledge and enthusiasm for his work inspired me during the struggling phases of the thesis. I would not have been able to complete the challenging part of the thesis without his help.

I would like to acknowledge helpful suggestions from my committee members: Dr. Ramon Lawrence, Dr. Patricia Lasserre and Dr. Jahangir Hosain.

I owe a huge gratitude to my loving family. Their love and confidence in me encouraged me throughout the course of thesis.
Dedication

I dedicate my thesis to my loving parents, Narayan Chandra Nath and Bandana Devi. I want to thank them for supporting me through thick and thin and standing like a rock beside me.
Chapter 1

Introduction

Public touch interactive devices are becoming predominant in public spaces for a wide range of activities. They allow people to perform various tasks by themselves using touchscreens. For instance, users can purchase transport tickets in transit stations, manage financial transactions in banks, or order food in restaurant kiosks. Additionally, they can self validate passports on arrival or print e-tickets from self-service terminals at the airport to name a few. Self-service terminals have also reduced the load on public administrators and led to efficient use of government funds and staff to serve more people [27].

Due to its convenience and intuitive ease of use, the touch modality is the primary input paradigm to interact with such public devices. In addition, there are other practical benefits for using touchscreens, such as requiring less space as both input and output are on the touchscreen. Traditionally, mechanical buttons were used as input to the devices. However, they tend to get damaged due to repetitive use and cannot easily be redesigned or mapped with new functions. In addition, mechanical buttons do not support advanced interactions such as mid-air gestures, which are commonly used on smart devices to increase input bandwidth with touchscreens [12, 70].

There is a growing concern that public touchscreens can become hubs for various health hazards - including infectious diseases - due to the large number of users interacting with them on a daily basis [98]. Additionally, with the Covid-19 outbreak, health experts have warned against touching public surfaces [110]. This warrants an exploration for alternative inputs to interact with public touch screen devices that do not require the users to come in contact with any surface. Touch-less interaction will also reduce costs as the terminal would require less sterilization [110]. In addition, there are existing practical design problems of touchscreen interfaces, such as the Fat Finger problem [111] which refers to the fingers being large for selecting and viewing items and becoming difficult, especially when items are close to each other.

Designing effective interfaces for public interactions is a challenging problem since different people of diverging capabilities, ages and, expectations
will interact with the public interfaces. We also have to explore input space for alternate interaction techniques for public touch interactive devices. Figure 1.1 highlights the palm, on-body and mid-air as possible input spaces. Nevertheless, many researchers have attempted to theorize or postulate how to design effective public interfaces [6]. Researchers have explored ways to shift the input space to users’ body locations. For instance, prior research demonstrated ways to use on-body surfaces (e.g., arms) as alternative input spaces to interact with mobile and wearable devices. The palm and arm areas can be used as both input and output medium with instrumentation such as using mini-projectors projecting from the top on the user’s hands [47]. Similarly, projectors and depth sensors can be placed on the shoulders for detecting gestures on hands [44]. In addition, researchers showed that our body surfaces could be used as an alternate place for gestural inputs such as scrolling or panning [48]. Arm space around smartwatches has been shown to be useful for accommodating virtual buttons [75]. In addition, researchers investigated ways to map smartphone UI elements and layouts onto users’ palms [42], even using the space on and above the back of the hand as an extended input space for smartwatches [116].

All these solutions focus on using on-body surfaces for interacting with personal devices such as smartphones or smartwatches and are not designed for public touch interactive devices. In addition, some of the prior work investigated in a context where users have to look directly at their palm. Such external instrumentation to switch the display space may not be practical for interacting with public touch interactive devices already equipped with
Figure 1.2: PalmSpace: an input technique leveraging users’ palms to interact with public devices.

Consequently, I envisioned a system where palms can be used as an alternate input space when facing towards the devices’ camera, keeping users focus on the existing device’s screen, without the need for additional hardware, and hence limiting the changes required to switch from touch interaction to palm interaction. To the best of my knowledge, leveraging the palm to mirror a touchscreen input space and to access public device functionalities has never been explored before. My research is also relevant in the context of current health risks associated with Covid-19. The prototype only requires an external camera, and users do not need to wear any additional devices that can allow my prototype to be easily deployed for public use.

In my thesis, I explored PalmSpace, that leverages the palm of a user’s hands as an alternative input space for public touchscreens interactions. This method of interaction is like Augmented Reality (AR) as the prototype is adding digital content on top of the users’ palms. I designed PalmSpace for allowing users to access on-device content using their palm as shown in Figure 1.2. With PalmSpace, UI elements are anchored on users’ non-dominant hands, mimicking the actual physical screen content. Users can use their dominant hand to select UI elements.

Figures 1.3 and 1.4 summarize the studies performed to design and eval-
uate the PalmSpace prototype. In order to design the PalmSpace, I first
surveyed to refine factors such as users’ preferences for using palm as an
input space, solutions to initialize UIs on the palm, or solutions to trigger
items selection. The survey results reveal that users prefer palm and mid-
air over on-body locations, dwell and depth-based tap selection triggers over
other hand gestures, and one-handed over two-handed initialization.

Based on the survey results, I conducted a pilot study exploring users’
preference for placing on-screen items onto the palm (Figure 1.2). The
participants were asked to rearrange buttons positioned in different arrange-
ments onto their palm and finger region while maintaining implicit relations,
such as order, among the targets. The initial arrangement of the buttons
represented the likely positions of those buttons in an on-screen interface.
From this design study, I identified the preferred palm and finger locations
for placing targets and how to fit the targets while maintaining the positional
information of the targets. Using the recommendations from the survey and
the design study, I developed a prototype, which implemented the preferred
interaction modes. The prototype used on-device cameras to track the palm
gestures of the participants.

I further investigated the palm space for interaction using the proto-
type with different design and evaluation studies. The two design studies
examined (i) appropriate item size and location on palm, (ii) suitable item
mapping from the touchscreen onto the palm, and (iii) suitable layout to
place the items. During the evaluation study, I asked participants to pre-
tend to purchase tickets via a custom transit system vending application.
Results reveal that participants were more efficient with PalmSpace than
with mid-air interaction. Finally, I installed PalmSpace in two public spaces
- in a restaurant and a university main hall - to gather valuable feedback
regarding the concept of using palm for interacting with public touchscreen
devices.

My contributions through the research work of my thesis are as follows:
1) PalmSpace, a novel interaction technique that leverages palm as an input
space for public touchscreens; 2) an evaluation of design factors related to
palm interaction, and 3) a demonstration of PalmSpace’s benefits via in-lab
and in-the-wild studies with an information exploration and decision-making
task in public space.
### Figure 1.3: Summary of design studies

**PART I: DESIGN**

<table>
<thead>
<tr>
<th>SURVEY (preferences)</th>
<th>PILOT (preferences)</th>
<th>STUDY 1 (performances)</th>
<th>STUDY 2 (performances)</th>
</tr>
</thead>
<tbody>
<tr>
<td>- Palm, mid-air, on-body</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Shared design factors</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Item placement</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Remediation</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Grid size</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Grid location</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Outcomes:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Palm and mid-air</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- One hand initiation</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Relaxed palm</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Outcomes:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- On palm</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Use landmarks</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Outcomes:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- 52H</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Up to 3x3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Palm</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Outcomes:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Any</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Left-to-Right</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Full presentation</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Figure 1.4: Summary of evaluation studies.

**PART II: EVALUATION**

<table>
<thead>
<tr>
<th>STUDY 3 (performances)</th>
<th>STUDY 4 (Discussion)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ticket vending scenario</td>
<td></td>
</tr>
<tr>
<td>- Palmspace vs mid-air</td>
<td></td>
</tr>
<tr>
<td>Outcomes:</td>
<td></td>
</tr>
<tr>
<td>- Completion time: Palmspace</td>
<td></td>
</tr>
<tr>
<td>- Preference: Palmspace</td>
<td></td>
</tr>
<tr>
<td>Outcomes:</td>
<td></td>
</tr>
<tr>
<td>- Positive subjective feedback</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>STUDY 4 (Discussion)</th>
</tr>
</thead>
<tbody>
<tr>
<td>In-vivo</td>
</tr>
<tr>
<td>- Main hall</td>
</tr>
<tr>
<td>- Restaurant</td>
</tr>
</tbody>
</table>
Chapter 2

Related Work

In this thesis, I examined the use of the palm space to facilitate tasks that require users to browse and interact with items beyond the conventional touchscreen space. This work is primarily inspired by previous research on mid-air and body-based interaction. In this chapter, I reviewed the previous work that investigated the aforementioned interaction spaces. I divided my reviewed research work for designing PalmSpace into two major input spaces, namely, Mid-Air and Body-based interactions. Each of these areas are further categorized under Input Techniques, Feedback Methods, Applications and Acceptance themes.

2.1 Mid-Air Interaction

2.1.1 Input Techniques

Panayiotis et al. have explored ways to extend touch input capabilities by shifting the input to mid-air space [71]. Researchers have investigated different sensing mechanisms that allow users to track their gestures in real-time in mid-air space. The following sections group common sensing technologies investigated in contemporary research endeavors.

Magnetic Sensors

Magnetic sensors have been demonstrated as promising solutions for detecting hand and finger positions around wearable devices [8, 86]. For instance, Harrison et al. [46] exhibited that finger states in mid-air could be tracked by wearing a magnetic ring on the fingers. The magnet (i.e., ring) causes disruptions unique to different gestures in the magnetic field within the vicinity of a multi-axis magnetometer, which was further used to detect different hand activities on the devices. In addition, Kent et al. [82] showed that variations in the magnetic field and prior knowledge of the hands’ expected kinematics could be used to determine the hands’ positions and rotations in 3D space. Hwang et al. [60] also explored using magnets to
write in mid-air space by attaching a magnet on the tip of a stylus which can be re-imagined for finger gestures by attaching a magnet on the fingertips.

**Infrared proximity sensor**

Infrared proximity sensors are commercially available and widely used by researchers for detecting the presence of fingers around smart devices (e.g., smartphones). These sensors usually come in pairs of emitters that transmit IR signals and receivers that detect signals reflected from nearby objects. Butler et al. [15] showed that infrared sensors could be embedded at the sides (i.e., edges) of smartphones to detect fingers’ activities around the device on either side of the phone to provide multi-touch input. In a similar work, Kartz et al. [72] demonstrated that hand gestures in the 3D space around a device can be detected by placing a set of IR sensors. Withana et al. [131] also investigated the efficient and optimized design for IR sensor placement (e.g., compact design, low energy consumption) while ensuring gesture detection accuracy.

**On-board sensors**

Inertial Measure Unit (IMU) is an electronic device that can measure different metrics such as its angular rate, acceleration, and orientation. IMU sensors can include miniature and embedded versions of sensors such as accelerometers, gyroscopes, and inertial sensors. Commercial wearable devices, such as smartwatches, can leverage IMU data to detect the gestures of the hand wearing the smartwatch [80, 126, 137]. Lu et al. [81] demonstrated that detecting dual-hand gestures are possible using two off-the-shelf wrist-worn devices (e.g., smartwatches) where both the devices detect the unique signal patterns for each hand during different hand-to-hand gestures. Similarly, accelerometer and gyroscope [74, 128], Photoplethysmography (PPG) signals from optical heart-rate sensors [139] have been shown to be promising solutions for detecting the movements of wrist and fingers [139]. Although a new interaction space can increase the possibility of a diverse set of gestures, only a few gestures can be detected conflict-free depending on the robustness of the prototypes [139].

Data from on-board sensors have also been used to locate the device position and orientation from the user’s body. For instance, Chen et al. [17] leveraged compass data to detect horizontal orientation and the accelerometer data to detect the vertical orientation of smartphones from the body. They also confirmed that the device location and orientation could
be derived from sensor data with high accuracy. Piezoelectric sensors use the piezoelectric effect to measure changes in pressure, acceleration, etc., by converting them to an electrical charge. These sensors have been shown to be useful to detect sound generated by different gestures using a set of microphones as input sources [134]. Off-the-shelf electric field sensing (EFS) chips, such as GestIC [89], can be used to track hands and fingers in the 3D space around devices within a small range of around 15 cm from the sensors [77]. Nandakumar et al. [96] have shown that the microphones and speakers of commercially available devices, such as smartphones and smartwatches, can be converted into active sonars to accurately detect the changes in finger movements in the mid-air space.

**External sensors**

Researchers have been using commercially available solutions, such as Vicon motion tracking system [121] and OptiTrack [102], Microsoft Kinect [130], the Leap Motion [78] and the RealSense cameras [61] for tracking users activity. These devices commonly use infrared sensors or multiple color cameras for stereo depth estimation.

**Electromagnetic Waves**

Different communication protocols such as WiFi, radio, and GSM signals use different ranges of electromagnetic field wavelengths as communication media. Interference in these electromagnetic field wavelengths can be used to track mid-air gestures around devices. There are commercial solutions that leverage the properties of electromagnetic fields for gesture detection. For example, Soli uses embedded Radar sensors and has a small form factor, and such devices are suitable for prototyping in design research [34]. GSM stands for Global System for Mobile communications and is used as a communication protocol for cellular networks. GSM uses 890 - 915 MHz [132]. Zhao et al. [140] have shown that GSM pulses can be used for recognizing hand gestures in mid-air.

**Vision-based sensing**

Robust gesture detection in mid-air is possible using only the RGB cameras of smartphones [37, 113]. Vision-based detection is also robust enough for practical applications, such as typing using mid-air gestures [99]. Wearable devices such as smartwatches are gradually becoming more versatile and incorporating many sensors such as RGB cameras, which are powerful
enough to implement different visual applications [120] and open doors for new interaction design possibilities.

Even though the pixel resolution has improved over the years, RGB cameras have a few limitations. For example, a single camera can only record 2D images of the 3D view, and the camera’s field of views are often limited due to fixed physical design (e.g., small form factor). Yang et al. [136] investigated unique ways to tackle the limitations, such as extending the field of view by using omnidirectional mirrors placed on top of the front camera of smartphones to enable 360° peripheral vision around the device. The panoramic images obtained using the mirrors are processed to enable mid-air gesture detection around the device in real-time. In addition, different vision-based algorithms are available to estimate the depth of objects in the input from a single camera. Machine learning models trained on real and synthetic data sets are used to estimate depth in real-time relative to object landmarks present in the input images from the camera [18, 87]. There are techniques in computer vision literature, such as stereo vision and time-of-flight, for estimating the depth of objects visible in multiple 2D images and they come packaged as commercial solutions [61, 78, 114, 130]. Commercial solutions can be used for exploring mid-air gestures in designated 3D spaces. For instance, researchers explored detecting mid-air gestures in front of smartphone devices by mounting a commercial depth-sensing camera [18, 73].

Furthermore, real-time depth sensing using vision-based cameras is possible in modern commercial phones and wearable devices. These camera systems usually implement stereo vision in which two or more cameras focus on the same scene, and the differences of angles and translations of the same objects in different scenes are used to estimate the depth of the objects. This information allows researchers to detect gestures in the 3D space covered by the camera’s view. Depth sensors, such as [114], can also be mounted on head-worn-display (HWD) to detect mid-air gestures as users interact in the augmented and virtual worlds.

2.1.2 Feedback Methods

The mid-air input offers unique interaction possibilities; however, such interactions require suitable feedback mechanisms that are expressive enough to present information related to users’ mid-air activities. Researchers have explored different visual and non-visual feedback mechanisms to convey users’ mid-air activities, which I will discuss next.
2.1. MID-AIR INTERACTION

Visual Feedback

Prior research explored different ways to visually present users’ mid-air activities around a smart device (e.g., smartphones). For instance, researchers [11, 49–51, 62] used on-screen feedback that visually presents users’ hand and finger movements on the device screen. They commonly dedicated a small space on the screen, showing a miniature view or overview of the entire interaction space and users’ hands/fingers on the overview with abstract shapes (e.g., a circle on the overview to represent a finger). Ens et al. [22] showed that Overview visualization is effective for displaying expanded work spaces on small devices.

Contextual cues have also been used to present off-screen items on the device’s screen. Researchers commonly use arrows or similar visual representation (e.g., wedge) on-the screen to show off-screen object locations and directions. For instance, Gustafson et al. [39] used on-screen wedges whose properties such as orientation and length depend on the off-screen items’ locations and directions from the device. Hossain et al. [59] designed EdgeSplit that shows off-screen items to the edge of the screen with circles. Burigat et al. [14] have compared approaches such as Wedge with the Overlay visualization and found that users were faster and more accurate accessing off-screen items using Wedge. In contrast, Overlay is more expressive for presenting information about objects [32].

Halo is another technique to convey spatial information in which a circle encompasses an object at its center and a small portion of the circle is displayed as an arc at the edge of the screen [10]. This arc shrinks as the user goes closer to the object in the off-screen space. With user studies, authors showed that the technique could be useful for visualizing off-screen content.

Visualizing off-screen content on small screen devices is challenging due to their limited display space. This can be mitigated by presenting off-screen items in their exact location beyond the device display space. While this objective can be difficult, emerging technologies such as holographic projection can be used for this purpose despite the requirement of bulky external projectors and limited display resolution. There are commercial solutions, such as VisBox [122], Epson Moverio [25] and Microsoft HoloLens [54], that can be used to view off-screen objects in the augmented and virtual space. Commercial head-worn displays (HWD) can be used to simulate many objects and interfaces itched inside an inner sphere around the user, similar to a cockpit [23, 24]. Commercial spectacles, such as Epson Moverio [25], have miniaturized projectors and reflectors within the lens, which can allow
users to view off-screen items in the augmented space observed through the spectacles [36]. An array of LEDs can be used to provide simple direction, and feedback from interactions and changes in the internal states of prototypes [29, 92, 103]. Prior research explored different parameters for output using LEDs, including intensity, direction, and distance of the light from the LEDs. Fog emitting from a fixed source (e.g., top of a table) has also been used in unique ways to provide feedback in mid-air [85, 105]. They used fog to create a semi-transparent interactive surface for reflecting projections, where the visual projections reflected from the fog changed based on the user interactions, which can be detected as gesture input [85]. Voxels, which are 3D versions of pixels in computer displays, can be represented as cubes suspended in mid-air and change the colors of the sides of the voxels to emit visual information in mid-air [109].

Non-Visual Feedback

Researchers have explored different ways to provide non-visual feedback on mid-air items [28, 43]. For instance, they explored providing tactile feedback using pressure created by synchronizing a large number of ultrasound transducers [55, 56, 58]. Others investigated vibrations as a channel to provide feedback for applications such as typing in mid-air space [99].

Prior research investigated the effects of location within the palm regions on the accuracy of haptic feedback by placing a grid of ultrasound transducers [129]. They found that users can accurately identify haptic feedback on the palm when the feedback is closer to the base of the palm. However, the accuracy did not vary when the users traversed their palms along the left to right direction.

An array of ultrasound transducers can be used for exploring the users' ability for discerning multiple haptic feedback in mid-air space [16]. The array of transducers was set up horizontally on the table, and the users interact with their hands directly above the transducers. The researchers found that users could identify feedback from ultrasound at different frequencies simultaneously and estimated the minimum distance required between two feedback locations for accurate identification. The transducers can also enable interesting applications such as emulating multi-finger touch input on an invisible horizontal screen in mid-air.

Bursts of air vortex can also be used for providing haptic feedback of impacts with moving objects in mid-air [112]. Sodhi et al. [112] designed a prototype that used a depth sensor to track user movements and provide air vortex generated by an actuated flexible nozzle to provide tactile feedback.
on the users’ hands. In a similar work, Gupta et al. [38] investigated the relevant factors such as vortex velocity, feedback delay, and the dimensions of the vortex generator for effective usage of air vortex for mid-air interaction.

2.1.3 Applications

The application use-case scenarios explored for mid-air space include accessing application features, workspace navigation, game or imaginary object controller or tangible interaction, and are discussed below:

**Accessing On-Device Item**

Researchers leveraged mid-air space for selecting on-screen items. For instance, Harrison et al. [46] showed that the space could be used for discrete item selection (e.g., buttons). In a similar work, Kratz et al. [72] shows that the mid-air space can be leveraged to highlight and select a color from a color palette. Mid-air space has been shown for menu selection, navigation and application management [113]. Researchers also demonstrated the space to be used for changing parameters within an application (e.g., scrolling through different widgets). For instance, they also explored mid-air space around the device for interacting different widgets used for scrolling through a discrete set of options, such as changing the song in a music player [69], and changing volume using a slider [136]. In addition, they also demonstrated that the space could be used for controlling an on-device cursor and making item selections [46] on small devices such as smartwatches with high accuracy. Researchers also exhibited that mid-air gestures could be used for responding to phone calls [67, 68] and that these features could allow users to interact in situations where the users cannot use their hands to interact with the touchscreen.

Mid-air space can also be used for multi-touch operations such as typing text on devices without the aid of touchscreens. For instance, SideSight [15] and Magic Desk [13] demonstrated that the mid-air space around devices can be used to support tasks that require multi-finger interactions, such as panning and zooming a map. Mid-air 3D space can also be useful for text or digit entry using a magnet to track the fingers involved with the tasks [66], instead of using an on-device keyboard or touch-pad. In addition, prior work demonstrated that text entry in mid-air space help to reduce screen occlusion issues [17, 66, 99].

Application shortcuts on mobile devices can be placed in mid-air space as demonstrated in Virtual shelves [79]. Here shortcuts were arranged in
2.1. MID-AIR INTERACTION

a grid and placed in a circular hemisphere in front of the user’s body. A user could trigger an application shortcut by moving the smartphone on the corresponding grid. With a user study, they demonstrated that virtual shelves were faster than the touch interface on the mobile device. Mid-air space can also be used as storage space for interactive objects which is examined in the ‘Pile Across Space’ concept [124]. Here, users could store objects in virtual piles around a device with flick gestures and retrieve the objects by moving the device to a corresponding spatial location.

Workspace Navigation

In addition to placing items, mid-air space can be used to navigate workspace as alternatives to the conventional flick and pinch gestures on touchscreen [51, 64, 115]. The mid-air space is larger than the conventional touchscreen, and prior research established that the space could be used efficiently for workspace navigation [64, 115]. For instance, Hasan et al. [51] demonstrated that mid-air space can also be used for navigating maps. They also confirmed that users are faster with navigating maps using mid-air gestures compared to the flick and pinch gestures on touchscreens. In addition to maps, standard documents (e.g., pdf or word files) can be navigated with mid-air gestures with the benefit that users can interact without occluding the screen [18]. Takashima et al. also showed that mid-air could be used in conjunction with on-screen space to enable map and document scrolling [118]. Additionally, different wearable devices, such as smartwatches and smartphones, can be paired to create unique possibilities of using mid-air space, such as map panning, where the smartphone and smartwatch locations in mid-air can be used to access map virtually placed in front of users [50].

2.1.4 Acceptance

While in-air interactions have been shown to be promising, a natural progression demands further investigation for exploring users’ acceptance of usage in public and private spaces (e.g., transportation, home). Consequently, researchers conducted studies examining acceptance of using mid-air gestures. For instance, researchers investigated social acceptance of mid-air interactions in public spaces [64, 73, 91, 107]. They commonly ask participants to use mid-air to interact with device content and then ask how they would feel using mid-air gestures in public, and private spaces [64]. In addition, they explored factors that can dictate the social acceptability of mid-air
2.2 BODY-BASED INTERACTION

gestures. For instance, Montero et al. [91] explored Suspenseful gestures, which referred to the extent users had to perform large gestures to trigger an action, and Magical gestures that referred to performing an action that seemed 'magical' such as toggling a light switch with smartphones. Results from a user study confirmed that people were opposed to gestures for which they had to perform large gestures noticeable by surrounding people. Researchers have explored the acceptability of hand gestures in the 3D space around portable devices such as a smartphone. For instance, Ahlstrm et al. [1] found that users decide to use gestures based on how the surrounding people perceive them. In addition, gesture properties such as smaller duration and gesture distance from the smartphone influence the acceptance of the gestures. While most research on the social acceptability of gestures focused on the point of view of the performer, researchers also explored social acceptability from the perspective of the observers who witness performers using gestures in different private/public context. For instance, Alallah et al. [3] explored the social acceptability of five input modalities, namely touch-pad, hand gestures, head gestures, voice commands, and ring, from both performers' and observers' perspectives. They found that the performers' perspectives of social acceptability do not always match the observers' perspectives which can even change based on input modalities. It should be noted that social acceptability data does not vary significantly whether the data is collected from laboratory experiments or through crowd sourcing [4].

2.2 Body-Based Interaction

2.2.1 Input Techniques

Previous work has explored body-based interaction, including touch or gestures on different body parts. The technologies used for sensing input on-body are similar to the sensing techniques for mid-air interactions with additional options such as wearable artificial skin for detecting touch gestures [9]. In general, on-body interaction is different from interacting with mid-air surfaces. Consequently, researchers have investigated possible factors that are important for designing on-skin interactions, such as gesture type and gesture location. They concluded that on-skin could support more expressive gestures than conventional touchscreens, and the forearm and the hand are the most preferred locations [127]. I grouped these previous research work based on the input techniques in the following sections.
2.2. BODY-BASED INTERACTION

Skin as Input Surface

In an early work, Harrison et al. [48] demonstrated the skin as a promising input surface for interacting with digital content projected on the user’s arm. Inspired by their results, researchers examined the use of the palm as an additional input space to trigger pre-defined functions [76], perform 3D rotation [73], or use it as an input space for augmenting keyboards to type on smart wearable devices such as Google Glass [123].

Researchers considered different techniques of developing artificial skin consisting of a stretchable conductive layer sandwiched between two insulating layers. This skin can detect the location of touch gestures by tracking the changes in resistance of the conductive materials due to deformations during touch gestures [57]. Such artificial skin can be complemented with detecting touch gestures by tracking the changes in the capacitance between the conductive layers, similar to touchscreens, and detecting touch gestures more accurately [57].

On-board and External Sensors

Accelerometers and gyroscopes on wearable devices, such as smartwatches, were used to identify hand and finger gestures on the skin around the device [135]. Unconventional input devices, such as stethoscopes which are commonly used for listening heart bits, were used to detect acoustic signals propagating through solid surfaces [45]. These acoustic signals are unique to different gestures and different surfaces, such as tabletop or walls, and accurate enough to detect different on-body scratching gestures [45].

Le Goc et al. [77] also designed a thin transparent prototype embedded with an array of commercial electric field sensing (EFS). Similarly, Zhou et al. [141] attached arrays of electric field sensing (EFS) to wearable devices, such as smartwatches, to enable electric field sensing on the body area around the prototype. This further was used to detect the device’s on-body hand and finger activities.

Infrared proximity sensor

In addition to detecting gestures in mid-air space, IR sensors can also be used with wearable devices such as smartwatches to track fingers in mid-air around devices and detect touch gestures in areas such as above the back of the hand [95]. Infrared LED-detector pairs can be attached to a smartwatch’s side facing the back of the hand for tracking the index finger of the other hand in a 2D space while using piezoelectric sensors for detecting
2.2. BODY-BASED INTERACTION

2.2.1 Touch Gestures

The back of the hand provides the sensation of tactile feedback for touch gestures and allows the user to interact without looking at the screen. Moreover, IR detection can be used anywhere on the body for tracking and providing tactile feedback.

2.2.2 Feedback Methods

The body areas such as the arm’s skin surface and the back of the hand can be used as an output medium. Similar to the feedback methods for mid-air interaction, feedback for on-body interaction can be divided on whether the feedback depends on visual sensation. The following sections categorize recent research on on-body interaction feedback into visual and non-visual feedback mechanisms.

Visual Feedback

The body surface has been used to provide feedback by augmenting different output techniques. For instance, smartwatches with pico-projector were used to project an interactive surface on the arm [75, 133], and electric field sensing was used to detect gestures on the arm around the smartwatch [141]. Harrison et al. [44] have also shown that the palm and other areas of the users’ body can be used for graphical multi-touch interaction, albeit requiring the user to wear a depth-sensing and projection system. Similarly, projector and camera systems can be used to transform any surface, including the arm and palm, as both augmented display and touch surface using markers on the interacting fingers [90]. Portable wearable projectors for providing visual feedback can be implemented using pico or laser projectors or an array of LEDs. For instance, laser-based micro-projectors have been used with cameras and mirrors to convert flat surfaces, such as tabletop, into both input and output medium [65]. Similar techniques have been used to convert on-body surfaces around wearable devices into output mediums [75].

Non-Visual Feedback

Researchers explored different non-visual feedback mechanisms to convey users on their on-body interaction. For instance, they used auditory feedback to provide information about the virtual item located on their palm [43]. Ueoka et al. [119] explored using bursts of air vortex to provide feedback for discrete actions, such as selections, on users’ faces to evaluate
the stress caused by such feedback. Their study results showed that feedback through air pressure can provide rich user experiences with negligible stress. However, the granularity of precise feedback using air pressure-based systems is lower than using ultrasound-based methods.

2.2.3 Applications

Sensors such as projectors and cameras can be attached to users’ bodies or in the users’ physical environment to enable palm, and body-based interaction [47]. Leveraging the system, Maity et al. [83] have considered pairing different application scenarios such as triggering communication or uploading a file using Electro-quasistatic Human Body Communication (EQS-HBC) transmission from an emitter worn on the arm. In addition, edge computing and AI inference for simple tasks are possible on artificial skin worn on the arm using off-the-shelf FPGAs, enabling real-time gesture detection on the body without requiring support from external hardware [9]. Gustafson et al. [41] investigated whether spatial knowledge of real-world interfaces (e.g., smartphone) could be used to make effective imaginary interfaces (e.g., Imaginary Phone). They used the common mobile widget designs and widgets’ relative positions to imitate a phone interface on the palm. They found that more than half of the participants relied on spatial memory to access their frequently used apps and transferred their spatial memory onto the imaginary interfaces. Moreover, their prototype with a depth camera demonstrated that participants could reliably interact with UI targets less than 17.7 mm in diameter, which mapped to a precision of 9.5 mm button sizes in the actual phone. The level of precision is enough to interact with most UI widgets on modern touch devices. Gustafson et al. [43] also found that users tend to depend on visually tracking their finger positions on the palm even when using imaginary interfaces. When visual cues are not available, such as when users are blindfolded, the imaginary interfaces can still be used with tactile feedback on the palm paired with spatial memory. Hence, spatial knowledge of existing user interfaces complemented with visual cues can be used to improve user performance for interacting with virtual mobile widgets on the palm. The same study found that users depend on tactile feedback from landmarks, such as the knuckle bones on the palm, when visual feedback of finger positions is not available to navigate the imaginary interface on the palm.
2.2.4 Acceptance

There has been limited research exploring the social acceptability of on-body gestures [107, 108]. Rico et al. [108] surveyed the social acceptability of various on-body gestures, such as touching the nose, tapping the cheek, or squeezing the forearm, in public settings and found user preference for gestures was contextual to the location and the audience around the users. During the studies, they displayed the participants’ videos demonstrating the gestures and surveyed the preference for using those gestures under different circumstances. Finger and palm spaces were revealed to be promising input spaces as a recent study showed that people were more comfortable interacting with their fingers than using in-air gestures [101].

Inspired by these promising results, I focused my investigation on leveraging the palm as a surface for interacting with public touchscreen devices. More specifically, I explored a novel palm-based interaction where a user has to put their non-dominant hand’s palm facing towards a device camera while interacting with the palm with the dominant hand’s finger. To the best of my knowledge, the use of palm in such a context has not been previously explored.
Chapter 3

Survey

I ran several interview sessions to collect user preferences on design factors for palm interaction with public devices. I was interested in investigating (i) how users want to interact with public devices while using the palm as an alternative to touchscreens (e.g., number of hands for interaction, UI initialization); (ii) how to display UI elements (e.g., remediation, anchor style, preferred location); and (iii) what are their preferred selection methods.

3.1 Participants

I recruited 33 participants (11 females, mean age 30.09 years, s.d. 12.24 with minimum 20 years and maximum 55 years) via emails, social networking websites, and word of mouth. Due to COVID-19 restrictions, I conducted online interviews via Zoom. All participants were right-handed and had prior experience using public touchscreen devices such as Bank ATMs (mean 5.61 years, s.d. 2.44).

3.2 Procedure

I met a group of 1-6 participants at a time using Zoom. I used a PowerPoint presentation and a narrative script to walk participants through my design space. The scenario started with a user walking up to a public touchscreen and initializing the device for interactions. Next, the scenario showcased different options for interacting with the device without using the touchscreen (e.g., selection trigger via mid-air tap or a shooting gesture, etc.). Participants could then access a Qualtrics [104] survey to select their preferences regarding each option. The presentation and questions order remained the same for each participant.

The Qualtrics form had four sections. The first section involved demographic information such as age, gender, and prior experience using public touchscreen devices. Figure 1.1 shows three possible interaction spaces -
3.2. PROCEDURE

Palm, On-Body and Mid-Air. In my thesis, I only focus on user preferences for the Palm interactions. These options comprised the Interaction Space which refers to the input area in motor space. The following sections describe the attributes for each of the interaction spaces:

3.2.1 Palm

The surface of the palm is investigated as an alternative to the touchscreen on the following factors:

- **Initialization** refers to how the prototype should detect a participant’s intent for interaction. The factors for initialization include (i) the number of hands (one-handed or two-handed); and (ii) the dynam- icity (posture or gesture) to initialize a window. Based on prior work [2, 21, 117], I included Open Palm, Closed Fist, Peace Sign, Relaxed Palm, Fist-to-Open-Palm and Two Hands Combination (Figure 3.1).

- **Remediation** There are several ways to provide visual feedback of augmented users’ interactions on a public touchscreen interface. For instance, the user interface can be shown with a superimposition of the palm. I refer to these situations as Hand-to-Screen (H2S) remediation (see Figure 3.2a).

Alternatively, I can virtually augment the user interface onto the user’s palm while showing the current camera feed to the user. I refer to this as Screen-to-Hand (S2H) remediation. (see Figure 3.2b). I can also have a Vignette remediation (see Figure 3.2c), i.e., a combination of S2H and H2S. With a Vignette, users see a superimposition of the palm on the screen (H2S), as well as a small live camera feed showing the S2H. For each remediation, I distinguish between two control-display ratios: the palm and the screen with the same height and the screen larger than the palm.

- **Anchoring** refers to whether the UI position is absolute (i.e., fixed in a given location), or relative (i.e., moves with the hand - see Figure 3.3 a-c). An Absolute anchoring can reduce jitters due to mid-air interaction fatigue but imposes a fixed location for interaction which can be uncomfortable. A Relative anchoring allows users to adjust the input space to a comfortable position but is prone to jitters due to chronic stress and unexpected movement of the anchoring hand.

- **Trigger** refers to actions to make a selection. For mid-air input, there have been several strategies proposed by industry and academy which
3.2. PROCEDURE

Figure 3.1: Dynamicity to initialize a window

- Open Palm
- Closed Fist
- Peace Sign
- Relaxed Palm
- Fist-to-Open-Palm
- Two Hands Combination

Figure 3.2: Remediation. (a) Hand-to-Screen, (b) Screen-to-Hand and (c) Vignette (combination of Hand-to-Screen and Screen-to-Hand). © Pinku Deb Nath, 2022.

can be categorized as triggering with the same hand as the anchor (Figure 3.4) or triggering with the hand that is not the anchor (Figure 3.5). Examples of mid-air triggers include pinch gesture with the
3.2. PROCEDURE

Microsoft Hololens [54], dwell [49], pinch [52], quick finger lift [49], or other variations [7] to name a few. Many of the triggers, such as pinch, tap, rubbing fingers, squeeze, wave, etc., can be detected using built-in gyroscope and accelerometers of commercial wearable devices such as smartwatches [128]. I included Dwell, Depth-tap, Double tap, Shooting gesture, and Pinch as potential selection triggers for palm input space (Figure 3.4 and Figure 3.5) as they are commonly investigated in prior work for different context (e.g., mid-air selection with Dwell [49]).

Different micro-gestures can be used to implement the aforementioned trigger gestures. The dwell trigger can be implemented by waiting on an item in the different input spaces [53], detecting the moment when the triggering index finger is lifted away from the desired target [106] which simulates the concept of lifting a finger after pressing on a physical button [106]. In addition to trigger methods, factors such as the size of the targets, input range, and target locations can also have different impacts on the aforementioned trigger gestures [49].

3.2.2 Results

The survey had different types of questions, including rank-order, multiple-choice and text-entry prompts. For the rank-order questions, the participants evaluated a number of options and rank-ordered them by dragging the options according to their preferences. The options with lower rank indicated that the participants preferred them more than the options with the higher ranks. For the multiple-choice questions, the participants responded with one choice for those questions with no alternative preferences. Finally, the text-entry questions are used to get qualitative feedback and observa-
3.2. PROCEDURE

Figure 3.4: Triggers with the hand that is the anchor for Palm input space: (a) Thumb gesture, (b) Open Hand.

Figure 3.5: Triggers with the hand that is not the anchor for Palm input space: (a) Depth-tap, (b) Shooting gesture, (c) Pinch, and (d) Dwell.

I reported the results of my analysis using Friedman and Wilcoxon tests and Bonferroni-corrected p-values for each factor: Initialization, Remediation (and the CD ratio), Anchoring, Trigger, and Interaction Space. Figure 3.6 summarizes the findings. Both Friedman and Wilcoxon tests are non-parametric tests. Friedman tests are suitable for testing factors with
ranked options similar to the responses in this survey and is a non-parametric alternative for parametric tests such as repeated measures one-way ANOVA.

Figure 3.6: Median user preferences for (a) number of hands and (b) gestures for initiating input space, (c) gestures for triggering selections and (d) overall preferences for input space. The lower values signify higher preference while the higher values signify lower preference.

**Initialization:** Participants prefer using one hand to initialize the interaction with the UI over two hands ($z = -5.40, p < .001, r = 0.88$). The parameters $z$, $p$ and $r$ are derived from Wilcoxon tests in which $z$ is statistic, $p$ is probability value that quantifies statistical significance and $r$ is effect size. I also found that the dynamicity (posture/gesture) has a significant effect on users’ preferences ($\chi^2(5, N = 33) = 122.28, p < 0.001$). Here, $N$ is sample size and $p$ is probability value for $\chi^2$ test. Relaxed and open palm are the most preferred solutions, with a mean ranking of 1.70 and 1.88, respectively. All pairwise comparisons are significant except between Relaxed and Open Palm, and Peace Sign and Close Fist.

**Remediation:** Remediation has a significant effect on users ranking ($\chi^2(2, N = 33) = 18.2, p < 0.001$): H2S (mean rank 1.49) and S2H (mean rank 1.97) are significantly preferred over Vignette (2.57). No other compar-
3.3 Discussion

Results suggest that one hand with a relaxed or open palm is the preferred solution to initialize interaction. Without a strong preference for using either H2S or S2H, I planned to rely on a follow-up investigation regarding performances to determine the best option. For the next steps and studies, I used relative anchoring (i.e., UIs move with the palm) as this was preferred by users over absolute anchoring. Participants ranked depth tap and double tap higher than other triggers. However, these triggers are difficult to detect accurately without a depth-sensing camera. Since I aimed for a solution requiring no additional instrumentation, I continued the rest of the studies with dwell as a selection trigger. Overall, the participants preferred mid-air over palm based on the survey. We believe that this is because participants are not familiar with interactions on the palm and it is a relatively new concept compared to interactions in mid-air.

In addition, I asked participants about privacy issues related to using Palmspace in public places. More specifically, I asked if they would feel comfortable seeing their faces on the public screens while using Palmspace. I included this question as previous studies [1, 91, 107, 108] showed that users’ willingness to use a new system depends on where they use it (e.g., public or private spaces). Participants said they would not feel uncomfortable if their faces were visible on the public display as long as the system did not record their actions.
Chapter 4

Design Exploration

When shifting on-screen content to users’ hands, UI elements can be placed on the palm or the fingers. Hence, I conducted a pilot study to investigate users’ preferences regarding the position of on-screen items mapped on their hands.

Figure 4.1: Combinations of on-screen items arranged in one group: (a) 2 items, horizontal layout, (b) 2 items, vertical layout, (c) 4 items, horizontal layout, (d) 4 items, vertical layout, (e) 4 items, grid layout with 2 rows, (f) 6 items, horizontal layout, (g) 6 items, vertical layout, (h) 6 items, grid layout with 2 rows, (i) 6 items, grid layout with 2 columns.
4.1 Pilot Study: Item Placement

In the pilot study, I investigated the following factors: *Number of items* and *Item arrangement*. I proposed that 2, 4, and 6 items be placed on the hand. The on-screen items are arranged horizontally, vertically, or in a grid layout. I considered single and double groups of items (e.g., one row of 4 items, two columns of 6 items, etc.). Note that, in the case of two groups, the groups of elements are aligned either vertically or horizontally. The items within one of the two groups have the same conditions applied to them as for items in a single group. Thus, there are 9 combinations of single groups (2 layouts for 2 items, 3 layouts for 4 items and 4 layouts for 6 items), and 18 combinations of double groups consisting of 4 layouts for 2 items, 6 layouts for 4 items and 8 layouts for 6 items. The combinations for a single group is shown in Figure 4.1 and the combinations of 2 items in each of the double groups are shown in Figure 4.2. Further examples of layouts for 4 and 6 items in double groups can be found in Figure A.1 and Figure A.2 of Appendix A.

![Figure 4.2](image.png)

Figure 4.2: Combinations of on-screen items arranged in two groups with 2 items per group: (a) 2 items with horizontal layout per group and the groups arranged horizontally, (b) 2 items with horizontal layout per group and the groups arranged vertically, (c) 2 items with vertical layout per group and the groups arranged horizontally, (d) 2 items with vertical layout per group and the groups arranged vertically.
4.1. PILOT STUDY: ITEM PLACEMENT

4.1.1 Participants

I recruited 6 participants among which two of them were females. The participants were aged between 20 and 31 years old. All of them had prior experience using public touchscreen devices. I conducted the study remotely via Zoom.

Figure 4.3: One group, horizontal (a, b, c) and vertical (d, e, f) arrangements of 6 items, (a, d) are initial states, (b, e) shows the result after a participant maps the items onto the palm image while (c, f) are the aggregated results of all the six participants’ input for the considered combinations.

4.1.2 Procedure

Before starting the session, I explained the study procedure and informed the goal of the study. Participants were provided with a PowerPoint file with one condition per slide. For example, Figure 4.3 (a) and (d) represents the initial states of the horizontal and vertical layouts of items in one group. Another example can be found in Figure A.3 of Appendix A which displays the initial states of items within one group in a 2-row and 3-row grid layouts. Each slide was composed of three parts: a screen with UI items, an unstructured cloud with the same items as the UI, and a drawing of a palm.
4.1. PILOT STUDY: ITEM PLACEMENT

Participants had to move UI elements from the unstructured cloud to the palm to showcase how they would mirror the screen UI.

Figure 4.4: Two groups representing horizontal group arrangement, 4 items arranged vertically within the two groups which are in horizontal orientation, (a) is the initial state, (b) shows the result after a participant maps the items onto the palm image while (c) are the aggregated results of all the six participants’ input for the considered combinations.

The slides were organized according to groups, with the first 9 slides for one group, while the rest of the slides for two groups. The items were labelled in alphabetical order (e.g., ‘A’, ‘B’... for one set, and ‘A1’, ‘A2’, ‘B1’, ‘B2’,... for two sets) where the suffix ‘1’ or ‘2’ represented which group the item belonged to in the original on-screen UI. In one group, the whole palm space is considered as a single region. There were 2 combinations for two items (items arranged horizontally or vertically), 3 combinations for four items (items arranged horizontally, vertically or 2×2 grid), and 4 combinations for six items (items arranged horizontally, vertically, 2×3 grid,
4.1. PILOT STUDY: ITEM PLACEMENT

or 3×2 grid).

Figure 4.3 (a, b, c) demonstrates the horizontal arrangement while Figure 4.3 (d, e, f) demonstrates vertical arrangements for one group with 6 items. In Figure 4.3 (a) shows the initial state of horizontal layout of items in one group and (d) shows the initial state of vertical layout of items in one group. Figure 4.3 (b) and (e) demonstrate samples of the preferences provided by one of the six participants while Figure 4.3 (c) and (f) show the aggregated preferences of all the participants for the respective combinations.

In two groups for item grouping, the palm space is divided into two regions which can be stacked vertically (side by side) or horizontally (top and bottom regions). There are 4 combinations for two items (2 palm group orientations combined with 2 item orientations which are horizontal and vertical), 6 combinations for four items (2 palm group orientations combined with 2 item orientations which are horizontal, vertical, and 2×2 grid), and 8 combinations for six items (2 palm group orientations combined with 2 item orientations which are horizontal, vertical, 2×3 grid, and 3×2 grid).

Figure 4.4 demonstrates the combination and result for 4 items arranged vertically within the two groups arranged horizontally. Figure 4.4a demonstrates the initial state of the slide representing the mentioned combination. The items layout at the top-right represents the on-screen layout for the group and items arrangements, and the bottom area has the copies of the items as proxies for the items of the on-screen layout. The participants relocated the proxy tags of items from the bottom-right area onto their desired locations in the palm image on the left side of Figure 4.4a. Figure 4.4b shows a preference of one of the participants with the blue and red-colored rectangular regions representing the preference of placing the groups with the same border color in the on-screen layout. Figure 4.4c shows the aggregated preferences of all the 6 participants for this combination. Another example of layouts for two groups can be found in Figure A.4 of Appendix A demonstrates the combination and result for 6 items arranged in grids of 2 rows within the two groups arranged vertically.

Participants were instructed first to touch their hands to enforce physical interaction before deciding on item locations. In addition, participants were asked to indicate whether they prefer to place items based on specific hand landmarks (e.g., joints) or based on the visual ordering of the on-screen items. I instructed participants to write down ‘impossible’ if they judged that a set of items could not be placed on the hand. I collected the modified PowerPoint files at the end of the session. A session lasted approximately 60 minutes.
4.1. PILOT STUDY: ITEM PLACEMENT

4.1.3 Results

I collected 162 slides (54 for single and 108 for double groups). Participants preferred to position items on the palm rather than on the finger area for single groups. In addition, participants tend to mirror the UI layout except if the group includes 6 items in a horizontal layout. Participants express that placing 6 items in a horizontal layout is complex and therefore prefer breaking their mirroring strategy to accommodate the items. For double groups, participants report 35 impossible combinations (16 using a horizontal layout and 19 using a vertical layout). Most combinations (27 out of 42 slides) are not possible when composed of 6 items in either horizontal (10 out of 12 slides), vertical (8 out of 12 slides), or in a grid layout (9 out of 24 slides). For both single and double groups, participants use landmarks for positioning items (115 out of 162 slides or 71%).

4.1.4 Discussion

From analyzing the aggregated preferences for each of the combinations examined in the study, we can observe a few patterns for placing items specific to different attributes of the combinations. These observations are discussed below:

- The participants preferred to use both the palm and finger areas, especially traversing along the fingers, to map items vertically when the items are arranged vertically within their groups in the on-screen presentation. I also observed that the participants preferred using palm and finger areas dedicated to items of the same group when the items are arranged horizontally or in a grid within the groups.

- Users tend to maintain the relative positional information of the items with their neighboring items if there is enough space in the palm and finger areas. However, when there is not enough space to maintain the original order of the items, for example, when placing 6 items horizontally, the participants tend to break the mirroring strategy but place remaining items in the nearby vicinity of the previous items. In addition, I observed that the participants tend to place items one at a time and do not tend to plan the mapping of items ahead of time.

- I observed that the participants preferred using landmarks on the hand, such as the joints at the base of the fingers, for all the combinations considered for items arranged in one group and 66% of the
4.1. PILOT STUDY: ITEM PLACEMENT

combinations considered for items arranged in two groups. I observed that when the number of items is small, users prefer the tips of the fingers, the base of the fingers, and the sides of the palm. It can be inferred that most of the participants preferred using landmarks on the hand to anchor items when the number of items is small and position items relative to previously placed items when the number of items is large.

− When two groups of items are stacked vertically on top of each other, the participants preferred placing items of the top group in the finger area and the items of the bottom group in the palm area. When the groups of items are arranged horizontally side by side, the participants preferred using the whole hand area for both the groups of items.
Chapter 5

Study 1

The survey revealed user preferences regarding the interaction initialization, anchoring style, and selection trigger. The pilot study for item placements revealed that participants could envision usage of both palm and fingers areas. The user preference for remediation techniques aggregated from the survey results was a balance between Screen-to-Hand (S2H) and Hand-to-Screen (H2S). There was no default choice for remediation technique based on user preference. Hence, I conducted this study to determine the suitable choice for remediation based on user performance. I also investigated combinations with other factors such as the number of items and the location of the input interface. I developed a prototype leveraging Augmented Reality techniques where I placed digital content on top of the users’ palms. To the best of my knowledge, no previous work related to hand or palm as an interaction space investigated factors that could influence users’ performance when using the hand space. Thus, I conducted my first study to quantify users’ performance on remediation techniques, the number of items that can be placed on the hand, and preferred items location (i.e., palm vs. hand).

5.1 Remediation Technique

Results from the survey did not reveal any clear preferences between Screen-to-Hand (S2H) and Hand-to-Screen (H2S) remediation techniques. Consequently, I investigated actual users’ performances to determine the appropriate choice for remediation, which I will use in the later studies. In the initial survey, I found that participants prefer using a virtual UI that covers the entire palm and also moves with the hand. Thus, the prototype dynamically moves and resizes the UI and its element according to users’ hand movements for both remediation techniques. For instance, the attached UI will also get larger if a user moves their hand close to the screen. However, only the on-screen representation of the palm-attached UI is larger, i.e., in the display space. UI elements remain the same size relative to the palm, i.e., in the motor space. With the S2H technique, users can directly
see themselves via the camera feed, along with the palm-attached UI on the non-dominant hand and a cursor controlled by the dominant hand to interact with the items (Figure 5.1, a and d). With the H2S technique, users can see an abstract hand representation along with the palm-attached UI (non-dominant hand) and a cursor (dominant hand) (Figure 5.1, b and c).

5.2 Grid Size

In order to represent the input space, I used a grid with targets arranged and identified on a row basis. I used different grid sizes for exploring the number of items that can be placed along with the horizontal and vertical directions on the user’s hand. I kept the grid size bounded inside the palm or the finger area. The pilot study for item placement revealed that placing 6 items on the palm or the finger area is difficult. I hence used a maximum
of 4 x 4 grid size - resulting in grids of 2 x 2, 3 x 3, and 4 x 4 items. Each cell (i.e., item) is visually scaled once the hand moves closer or further from the camera. However, the actual interaction space for the grid remains constant with respect to the size of the participant’s palms.

5.3 Grid Location

The input grid can be located either in the palm or the finger area. The palm area covers the space situated just under the fingers and over the wrist. The finger area covers the space from the start of the fingers up to the fingertips. I used the same grid size in both palm and fingers areas for consistency, i.e., approximately the palm width and height. The decision to investigate this factor was motivated by the analysis of the item placement study in which I observed that the participants tend to group related items of the same group or area in either the palm area or along the fingers.

5.4 Participants

I recruited 12 right-handed participants (4 females, mean age 26.83, s.d. 5.48 years with minimum 20 years and maximum 42 years) via on-campus flyers and word-of-mouth. All participants had prior experience using public touchscreen devices and received $15 for their participation. Due to restrictions regarding in-person studies, I conducted the study remotely via Zoom. I provided participants with a link to access the web application. The application ran on participants’ web browsers, received the image frames from their device camera, and leveraged MediaPipe’s tracking algorithms to detect hand movements. The web application sent trial-related data (e.g., trial time) to the database on trial completion.

5.5 Apparatus

Based on the design choices observed from the survey results and the pilot study for item placement, I developed a web application that uses participants’ device cameras to track their hand and finger movements. We used the information to augment a grid and a cursor. The input grid is mapped onto the left palm and the cursor is controlled with the tip of the right hand’s index finger. We used a grey-colored 10-pixel circle (0.25 cm) to represent the cursor. The web application was motivated by the portability and low cost since I am investigating an alternative interface for public
interface that can be integrated at a large scale. Moreover, I had to perform
the user studies remotely and use as few additional expensive hardware as
possible. Figure 5.2 demonstrates an overview of the overall system.

5.5.1 System Design

The prototype follows a client-server model in which a web application running in the participant’s browser is a client while the remote server serving the application and saving study data is the server. The following sections elaborate the components further.
5.5. APPARATUS

Server

The back-end was served using NodeJS [100] with ExpressJS [26] for handling routes. For storing data from the user studies, I used MariaDB [84] which is an open-source version relational database having the same functionalities as MySQL [94]. Requests to the server do not directly go to the Node [100] server. Instead, requests are first received by Nginx [97] which implements reverse proxy and TLS/SSL [31] certification. TLS/SSL are application layer protocols in the TCP/IP model that provides asymmetric key encryption from both client and server endpoints for providing secure connection. TLS 1.3 is the latest version while all the SSL versions are deprecated. TLS/SSL certificates are used to verify the ownership of a public key. The NodeJS server sends its certificate to the client upon initiating connection and the client verifies the certificate using its own Certificate Authority (CA) certificates. This additional step was required so that I can use Hypertext transfer protocol secure (HTTPS) for connecting with the server which is required by the remote server’s default firewall. HTTPS requests are received at port 443 and routed to port 80, which is listened to by the Node server. Both the NodeJS server and the database are hosted in a private server provided by the University of British Columbia.

Hand-Tracking

The front-end uses hand tracking models provided by MediaPipe [87, 88], a framework for developing applications that use trained classifier models and compiling those applications targeting different operating systems and environments, including desktop, mobile, and web browsers. The trained models are TFLite [35] binaries and can include models for detecting different body parts such as faces, hands, eyes, etc., from video streams [138]. The models are written in C++ and compiled to WebAssembly (WASM) [125] using the Bazel build system [33].

Web Application

The web application’s front-end was developed using JavaScript, HTML, and CSS. Most of the application logic are executed in real-time in the participant’s browser. The web application uses WASM binaries compiled using MediaPipe and hosted in the server. The compiled WASM binaries are fetched by the web application using ⟨script⟩ tags.

The web application accesses the video frames from the webcam and passes them to an instance of the hand-detection model in the JavaScript
run-time of the participant’s browser. The hand-detection model is instan-
tiated using the WASM binaries. The model returns sets of landmarks for
each input image frame, with each set belonging to one of the detected hands
in the image frame. Each set has different attributes and their values, such
as the confidence of detection and whether the hand is left or right. Each
set also has a list of 21 landmarks describing the locations of the hands in
the image frame as shown in Figure 5.3. Each landmark is a tuple of x, y,
and z coordinates scaled between 0 and 1. We need to multiply x and y with
the width and height of the image frame to get the actual coordinates. The
z coordinate is the relative depth or distance of the landmark location from
the camera relative to the landmark at the base of the palm of the same
hand.

![Diagram of hand landmarks](image)

Figure 5.3: 21 indices of hand landmarks starting from 0.

Once the sets of landmarks and other attributes from the current im-
age frame are available, the app triggers a callback function that processes
and updates the application state. First, the app checks whether and where
in the image frame to show the interface and the dimensions of the inter-
face. Next, the landmark coordinates are used to update the location of
the interface. Then the app updates the location of the cursor based on the
landmarks of the index finger of the right hand, which is the landmark at
index 8 of the set of landmarks as shown in Figure 5.3.
5.6. Procedure

Afterwards, the app determines whether a selection was triggered based on the current landmarks and the app’s previous state. For the prototype, I used Dwell as the trigger. Hence, the app determines whether the cursor is inside the area of any interactive item, how long the cursor has been waiting inside the area of the currently highlighted item and whether the wait time is greater than a threshold for the Dwell trigger. If the wait time exceeds the threshold, the app triggers a selection. Finally, the app renders the image frame, the interface, and the cursor onto the HTML Canvas using the 2D context. The context of HTML canvas maintains the internal state of the canvas and is maintained by the browser. Another option for the context is WebGL2, which is faster as it allows the use of GPU processors using the WebGL interface. However, WebGL2 and GPU support may not be available to all the participants, and using ‘2d’ context provided acceptable performance.

5.6 Procedure

During the study, participants were asked to stay connected via Zoom. They were then given the link to the web application, and the researcher ensured that participants selected the correct experimental conditions in the app options through screen share.

Participants raised their non-dominant hand and faced it towards the camera to start interacting with the system. Then, the web application displayed a start button and a grid of items on the hand. The start button was always placed to the left-center, outside the grid (Figure 5.1, a-d). To begin a trial, participants moved the cursor (index on the dominant hand) on top of the start button and performed a Dwell action to validate their selection. Next, the app highlighted the random trial target cell with a red background. Participants then moved the cursor on top of the target cell to complete the trial. The system highlighted hovered cells with a blue color. Previous research found dwell time of 662 ms performed the best [63]. Hence, I defined a dwell time of 600 ms to trigger a selection. Once the trial was completed, the system highlighted the ‘start’ button again to launch the next trial.

I used a $2 \times 2 \times 3$ within-subject design, with Remediation (H2S and S2H), Grid Location (Palm, Finger), and Grid Size ($2 \times 2, 3 \times 3, 4 \times 4$) as factors. Participants had to select every cell 3 times, resulting in 12 selections for the $2 \times 2$ Grid Size, 27 selections for the $3 \times 3$ Grid Size, and 48 selections for the $4 \times 4$ Grid Size. Thus, each participant had a total of 87 selections for
each Remediation × Grid Location combinations, for a total of 348 trials per participants. Participants were given practice trials to get familiarized with the techniques.

The presentation order of the Remediation was counterbalanced across participants. The order of Grid Location and Grid Size were randomized for each Remediation. Trials ended on successful selection. Participants were instructed to complete the trials as quickly and accurately as possible. I collected participants' preferences at the end of the session. Participation lasted approximately 60 minutes (including breaks and practice).

5.7 Results

I broke down the trial completion time (from trial start to the correct selection) into two parts: (i) Target visit time, from the trial start to the final selection attempt, and (ii) Selection time, the dwell time. As the selection can be made with other trigger options (e.g., depth tap), I did my data analysis on Target visit time. I used repeated measures ANOVA and post-hoc pairwise comparisons (Bonferroni α-level = 0.05) to analyze Target visit time, and Friedman and Wilcoxon tests with Bonferroni correction to analyze user preferences. The repeated measures ANOVA was used to analyze changes in mean Target visit time under the different factor combinations. The α-level of Bonferroni for post-hoc pairwise comparisons control the significance cut off threshold.

Target visit time (Figure 5.4): Grid Location has a significant impact on Target visit time ($F_{1,10} = 9.32, p < 0.05, \eta^2 = 0.48$) where F is test statistic, p is the p-value indicating significance and $\eta^2$ is the square of the correlation or difference ratio. Participants are significantly faster in the Palm area (mean 1997 ms) than in the Finger area (mean 2525 ms). I also observed a significant effect of Grid Size on Target visit time ($F_{2,20} = 17.68, p < 0.001, \eta^2 = 0.64$): The 4 × 4 (mean 2495 ms) grid is significantly slower than both the 3 × 3 (mean 2100 ms) and the 2 × 2 (mean 1601 ms) grids. There is no significant effect of Remediation ($F_{1,10} = 1.36, p = 0.27, \eta^2 = 0.12$). Participants are generally faster with S2H (mean 2235 ms) than with H2S (mean 2270 ms). I also observed that participants take longer time if targets are located near the little finger area (Figure 5.5, a-c), i.e., the furthest from the 'start' button. Additionally, this result confirms findings from previous work, showing that pointing performances degrade as fingers move away from a reference point when indirect feedback about the finger position is provided [40].
5.7. RESULTS

Figure 5.4: Mean target visit time for Palm and Finger across different Grid Size. Error bars: ± 2 S.E.

Figure 5.5: Target visit time in (a) 2×2, (b) 3×3 and (c) 4×4 grid cells.

Preference rating: I collected participants’ preferences using 5-point Likert scale questions. Participants rated four Remediation × Grid Loca-
5.8. DISCUSSION

I found differences between the four combinations ($\chi^2(3, N = 12) = 25.49$, $p < .0001$): Between H2S-palm and S2H-palm, H2S-Finger and S2H-Palm, and H2S-Finger and S2H-finger where $N$ is sample size and $p$ is p-value indicating significance. S2H-Palm is the most preferred technique (mean rating 4.75), followed by S2H-finger (mean rating 3.83), H2S-Palm (mean rating 3.08), and H2S-Finger (mean rating 2.08). I also observed significant differences among the three Grid Size ($\chi^2(2, N = 12) = 21.83$, $p < .0001$). I found differences between all pairs where $2 \times 2$ is the most preferred grid size (mean rating 4.92), followed by $3 \times 3$ (mean rating 3.67) and $4 \times 4$ (mean rating 1.83).

5.8 Discussion

The study results indicate that using the palm area is faster than using the finger area. Target visit time increases significantly for palm with the $4 \times 4$ grid size as many items have to be accommodated on a small space. Thus, I suggested avoiding using this grid size on the palm. In addition, item selection was slower toward the pinky’s side. Indeed, besides longer movement times from start to target, the hand detector seldom failed to detect both hands in case of strong overlap. Thus, items further away should be larger to accommodate both complications. In addition, based on observing the participants interacting with the interface, hand sizes did not affect the selection performance. However, environment factors such as varying light exposure affected the hardware performance. I did not see any significant difference on the Remediation. However, participants expressed their preference for using S2H-Palm and S2H-finger. Thus, I continued my exploration with these two options.
Chapter 6

Study 2

Study 1 shows that more than three items in a row affects selection time. However, I surveyed UIs for 8 public touch interactive devices (4 ATM, 3 ticket vending systems, and 1 food ordering self-checkout system) and observed that several UIs include four or five items in a row or a column (e.g., CIBC ATM [19], or City Bank ATM [20]). Hence, further exploration was needed to map more than three items from the on-screen UI onto the palm.

6.1 Design Factors

I explored solutions to map five items onto the palm area in this study. Based on the observations from the pilot study for item placements for situations in which there were 4 or 6 buttons one after another, I considered the following design factors for investigation in this study:

6.1.1 Layout

I considered two layout strategies: Grid and Flow. With the Grid (Figure 6.2, b), the system positions up to 3 items in the first row, and continue placing remaining items in rows below. The Flow strategy (Figure 6.2, a) is similar to writing on a piece of paper without correctly planning the space needed and continuing writing vertically near the edge.

For instance, this usually happens when the users do not plan correctly the necessary space needed to write the texts. The system positions up to 3 items in the first row, and continue placing remaining items in the last column as visualized in Figure 6.1.

6.1.2 Scanning Direction

The UI items can be scanned from Left-to-Right (Figure 6.2, a), i.e., the Scanning Direction in our geographic area, or Right-to-Left (Figure 6.2, b), i.e. the movement direction of the pointing finger. Indeed, for right-handed
6.1. DESIGN FACTORS

Figure 6.1: Flow layout concept with respect to scanning direction. (a, b, c) Flow layout with left to right scanning direction, (a) Original arrangement of targets (A-E) in on-screen UI mapped onto the palm, (b) Rearrangement of the targets according to Flow layout with C, D, E in the last column and (c) Rearrangement of the targets according to Grid layout with B, C, D, E having the same dimensions and D starting from the second row. (d, e, f) Flow layout concept with right to left scanning direction, (d) Original arrangement of targets (A-E) in on-screen UI mapped onto the palm, (e) Rearrangement of the targets according to Flow layout with C, D, E in the last row and (f) Rearrangement of the targets according to Grid layout with A, B, D, E having the same dimensions, A and B starting from the second row with C occupying the whole first column.

users, the right index finger enters the palm via the right side, creating a right-to-left scan of items.
6.1. DESIGN FACTORS

6.1.3 Integration

I explored two options to integrate PalmSpace in an existing interface: Full and Partial. In the Full version, the original interface has to be modified to let users see themselves via Augmented Reality on the full screen (similar to S2H in my previous studies). In the Partial version, a static palm icon
6.1. DESIGN FACTORS

Figure 6.3: Partial integration: (a) Left-to-Right with Flow, (b) Right-to-Left with Grid. © Pinku Deb Nath, 2022.

with the palm-attached UI is positioned in an empty space of the original UI. I also include the raw camera feed without any visual augmentation. I was primarily interested in finding the degree of effort the users would require to adjust with the PalmSpace interface. Full integration uses the PalmSpace interface only without showing the on-screen UI, while Partial integration
simulates the situation in which feedback from using the PalmSpace interface
is overlapped on top of the existing on-screen UI.

6.2 Participants and Apparatus

I recruited 12 right-handed participants (4 females, mean age 25.67, SD
5.92 with minimum 20 years and maximum 42 years) via flyers and word-of-mouth. All participants had prior experience using public touchscreen
devices (mean 9.17 years) and received $15 for their participation. The web
application was similar to the one used in study 1.

6.3 Procedure

I conducted the study remotely via Zoom. Participants were first given
information about the study and provided the web application link. They
were asked to share their device screen to ensure they selected the experi-
mental options correctly on the web application.

The task consisted in selecting an item from a set of items displayed
at the top of the screen (Figure 6.2, a). The participants were asked to
raise their non-dominant hand and face it toward the camera. Then, a start
button appeared near the thumb of the non-dominant hand. The system
indicated the target item with a purple rectangular border on the original UI
and a textual prompt (e.g., Please select apple). Participants used the live
augmented camera feed with Full or the icon image with Partial to locate
the item and to select it by positioning the index finger in the corresponding
area for 600 ms. Similar to study 1, a cursor indicated the index position.
A correctly performed selection ended the trial and showed the next trial.

6.4 Design

I used a 2×2×2 within-subject design with Layout (Grid, Flow), Scan-
ing Direction (Left-to-Right, Right-to-Left), and Integration (Full and Par-
tial) as factors. The order for Integration was counter-balanced across par-
ticipants. Layout and Scanning Direction were randomized across participants for each Integration. Each trial involved a target randomly selected
among available items. Participants performed 20 repetitions of each Lay-
out - Scanning Direction - Integration combinations, for a total of 160 trials
per participant. I used eight different sets of grocery icons - one for each
combination condition (Figure 6.2 a-b, Figure 6.3 c) - to prevent learning and memorizing items position across conditions. Participants completed 5 random practice trials before starting a new combination. I also collected subjective feedback via 5-point Likert scale questions. A session lasted approximately 45 minutes, including short breaks and practice.

6.4.1 Results

I removed 39 (2.03%) trials with completion times exceeding the mean time by more than 3 standard deviations.

6.4.2 Target visit time

I define the Target visit time (Figure 6.4, a) as the elapsed time between the trial start and the final attempt to make a successful selection, i.e., just before the dwell time action. I use repeated measures ANOVA and Bonferroni adjusted (α-level = 0.05) post-hoc pairwise comparisons. Integration has a significant effect on target visit time ($F_{1,11} = 28.36, p < 0.001, \eta^2 = 0.72$) where $F$ is test statistic, $p$ is the p-value indicating significance and $\eta^2$ is the square of the correlation or difference ratio. Full Integration (mean 2,169 ms) is significantly faster than Partial Integration (mean 2,963). I did not find main effects of Layout (mean target visit time for Grid 2,705 ms and Flow 2,428 ms) or Scanning Direction (mean target visit time for LtoR 2,416 ms and RtoL 2,717 ms).

Figure 6.4: (a) Mean target visit time and (b) mean error rates. Error bars: ± 2 S.E. (c) Median user ratings
6.4.3 Error rate

(Figure 6.4, b): The error rate corresponds to the number of trials with at least one selection attempt in another cell than the target, divided by the total number of trials. I use Friedman tests with Wilcoxon tests for post-hoc pairwise comparisons to analyze error rates. The overall error rate is 6.43% (121 trials with error). Integration has an significant effect on error rate ($\chi^2(1, N = 12) = 12.00, p < 0.001$) where $N$ is sample size and $p$ is p-value indicating significance. Pairwise comparisons show that Full (mean 4%) has significantly fewer errors than Partial (mean 9%) ($z = -3.05, p < .05, r = 0.72$) where $z$ is statistic, $p$ is probability value that quantifies statistical significance and $r$ is effect size. There is no significant difference in error rates regarding Layout and Scanning Direction options.

6.5 Preferences

(Figure 6.4, c): There is a significant effect of Integration on users preference ($z = -2.05, p < .05, r = 0.59$): Full (mean rating 4.42) is preferred over Partial (3.08). I also observe a significant effect of Scanning Direction ($z = -2.57, p < 0.05, r = 0.74$), with participants rating LtoR (mean rating 4.50) higher than RtoL (mean rating 3.33). I didn’t find any significant difference between Layout options ($z = -0.91, p = 0.36, r = 0.26$).

6.6 Summary

Results reveal that Full Integration has a better overall performance across target visit time, error rate, and preference. Hence, I decided to continue my following studies with this Full Integration. Though I did not see any significant difference of Scanning Direction on users’ performance, I saw a strong preference for using LtoR than RtoL Scanning Direction. I believe this is because all participants were native left-to-right readers. Indeed, participants commonly express a strong preference for tasks with a left-to-right direction than right-to-left [30]. The Scanning Direction is hence more important than the motor actions flow used by PalmSpace.
Chapter 7

Study 3

The previous studies provide us with information on the factors that could influence the performance of PalmSpace. Based on the study results, I selected the best values for the factors to design PalmSpace. I run a further study comparing PalmSpace with the mid-air interaction technique. More specifically, I accommodated different UIs commonly seen on self-service ticketing vending machines to compare the performance of the two techniques in ticket purchase tasks. I used the Dwell trigger for both the palm and mid-air input spaces instead of touch input since the dwell trigger had the same dynamics in both spaces. In contrast, touch input had different factors to consider in the two input spaces. For example, touching the physical surface of the palm can provide tactile feedback of touching a surface while touching in mid-air space is more abstract, and there is no sensation of touch. In addition, there are many technical parameters for touch input, such as depths of both the hands, ambient lighting, etc. Since I aimed to compare input spaces, I used Dwell trigger to reduce factors unrelated to the spaces that can influence user performance.

7.1 Participants and Apparatus

I recruited 12 right-handed participants (4 females, mean age 26.0 years, st dev 6.26 years with minimum 20 years and maximum 42 years). All participants had prior experience using public touchscreen interfaces (mean 8 years of experience). Participants were compensated with $15 for their time. I used the same platforms (e.g., web application using MediaPipe [87]) as previous studies to design the web application for accommodating UIs of the ticketing vending machines. I leveraged MediaPipe’s algorithms to detect hand movements captured via the participants’ device camera for both palm and mid-air interactions.
7.2 Procedure

As for previous studies, I met participants via Zoom to provide information about the study and the application link and ensure they ran the study properly via screen sharing. Participants were required to perform selection tasks to purchase train tickets using both palm and mid-air inputs. I designed a set of four UIs (i.e., Main UI, Ticket Type, Payment Amount, and Payment Method) to mimic a standard ticketing vending machine applica-
7.2. PROCEDURE

The current trial instruction (e.g., *Buy Ticket for One Way Travel costing $20 using Credit Card*) is placed at the top of all UI screens throughout the trial. A trial starts with selecting the start button, which also triggers a timer and displays the contents of the Main UI. The Main UI contains options, such as Buy Ticket, Buy Card, and Load to a card (Figure 7.1, a). Selecting an option from the Main UI opens the Ticket Type window to purchase a Day Pass, One-Way, or Two-Way tickets (Figure 7.1, b). After the option selection, the application navigates users to the Payment amount with options $10, $20, $50, $100 and $200 (Figure 7.1, c), and finally to the Payment method (i.e., Credit card and Debit card) UIs (Figure 7.1, d). If the participant makes a wrong selection (e.g., select $10 instead of $20 in the aforementioned example), the application ends the trial while marking it as an error and prompts participants to redo the trial straight away. Sequentially selecting correct items from all UIs ends the trial, stops the timer and shows the prompt for the next trial information.

Figure 7.2: UIs for Mid-Air Input Space, (a) Main, (b) Ticket Type, (c) Payment Amount, (d) Payment Method
7.3. DESIGN

With PalmSpace, users were required to raise their non-dominant hand and face it toward the camera. Once the application detected the hand, it displayed the UI onto the palm via the embedded video feed. Similar to the previous studies, a grey-colored circular cursor of radius 10 pixels (0.25 cm) was mapped to users’ right-hand index fingertip. Participants selected a UI element by hovering the cursor over the palm’s corresponding area for a predefined dwell time. Note that the presentation of UI elements on the palm was different for each UI (Figure 7.1) as I followed best practices for arranging the items as identified in the previous studies. For instance, with a UI containing more than 3 elements, I used the Flow layout to accommodate elements on the palm. Another example is the increased size allocated to elements further away from the start button to assist in faster selections (Figure 7.1, c). With Mid-Air, the UIs were placed in a fixed location at the center of the screen (Figure 7.2). I used a button dimension of 160 × 80 pixels (4.2 × 2.1 cm) in a 22-inch display. Similar to PalmSpace, a cursor was mapped to users’ right-hand index fingertip. Participants selected the UI elements by moving the cursor on top of an element and hovering over for a predefined dwell time. While using 600 ms for dwell time in the previous studies, I observed few errors related to unintended selections. Hence, I set dwell time to 1000 ms in this study in order to mitigate the errors in dwell selections. The grid containing the options in all the UIs has the dimension of 213 × 179 pixels (5.6 × 4.7 cm) in a 22 inch display. I kept the grid dimensions the same for both PalmSpace and Mid-Air to reduce the effects of different sizes of input spaces and distance traveled by the index finger.

7.3 Design

I used a within-subject design with Technique (PalmSpace, Mid-Air) counterbalanced across participants. The study application automatically created the instruction for each trial by randomly selecting an option from each UI. After 10 random practice trials, participants performed 30 repetitions with each Technique, for a total of 60 trials per participant. After completing all trials, participants were asked to provide subjective feedback on the techniques using 5-point Likert scale questions. A session lasted approximately 40 minutes, including short breaks and practice.
7.3. DESIGN

Figure 7.3: Mean (a) trial time and (b) cursor movement across two techniques. Error bars: ± 2 S.E. for Mid-Air and PalmSpace

Figure 7.4: (a) Mean error rate across two techniques with error bars: ± 2 S.E. (b) Median preference rating for Mid-Air and PalmSpace
7.4 Results

7.4.1 Trial Time

I used a t-test to analyze the trial time. I found a significant effect of Technique on trial time ($t(11) = 3.04, p < 0.05$) where $p$ is the p-value denoting significance for t-test. PalmSpace (mean time 11,295 ms) is significantly faster than Mid-Air (mean time 14,170 ms) (Figure 7.3, a).

7.4.2 Error Rate

I further analyzed the error rate defined by the number of error trials divided by the total number of trials (Figure 7.3, b). A Technique had no effect on error rate (Wilcoxon $Z = -0.98, p = 0.32$, Palm 15% and Mid-Air 22% across 5 interfaces) where $z$ is statistic and $p$ is p-value indicating significance.

7.4.3 Cursor Movement

During a trial, I also recorded on-screen cursor movement length to determine finger movement by participants for both techniques (Figure 7.4, a). A t-test showed a significant difference in the cursor movement length for Technique ($t(11) = 6.10, p < 0.001$). Participants moved the cursor more on the screen with Mid-Air (2,670 px) than with Palm (1,378 px). Participants rated PalmSpace (mean 4.33) significantly higher than Mid-Air (mean 3.00) (Wilcoxon $Z = -2.00, p < 0.05$) (Figure 7.4, b).

7.4.4 Preference

I processed and analyzed the post-survey feedback from the participants to determine the overall preferences of the participants. Preferences for the Palm and Mid-Air input spaces are acquired using 5-point Likert scales. A score of 5 for an input space means that this input space is most preferred, while a score of 1 means that the space is least preferred. Overall, the Palm input space with the average score of 3.71 (s.d. 1.41) is preferred more than the Mid-Air input space with an average score of 2.43 (s.d. of 1.37).

7.5 Summary

This study shows that, compared to the Mid-Air technique, PalmSpace interactions significantly reduce trial completion time. Hence, although the
participants preferred mid-air over palm in the initial survey conducted during the exploration phase, the performance of PalmSpace was better than interacting in mid-air. We believe that this is primarily due to the following reasons: First, with PalmSpace, items can be accessed easily by moving the index finger within a small palm area. Second, users can use haptic feedback on the palm which is missing in the mid-air technique. Third, users can leverage their spatial memory to access items located on the palm quickly. Hence, the palm is an efficient area to accommodate UIs, which does not require large finger movements for accessing on-screen items. In addition, Palm offers haptic feedback, which some participants found to be useful for mapping and accessing the items. Although the results suggest Palm is an efficient space for accessing on-screen content, I identified some limitations which I will discuss in Chapter 9.
Chapter 8

In The Wild Demonstration

The previous user studies have been conducted in a supervised environment for certain tasks (e.g., ticket purchase) in which I instructed the participants on how to use PalmSpace. Also, user studies were conducted remotely where participants used the system from their homes. We can envision PalmSpace can also be used in public space with minimum supervisions. Consequently, I set up PalmSpace in actual public spaces to collect feedback and observations in ecologically valid settings.

8.1 Conditions

I tested PalmSpace in a university building’s main entrance and in a restaurant. The setup consisted of a regular Dell XPS laptop with a WiFi connection, a table, and pieces of paper with “TRY ME” inscriptions (Figure 8.1). I included inscriptions since there was no interactive system to replace, and perhaps a laptop alone would not have had engaged passers-by. The main screen consisted of textual inscriptions “Show your left palm toward the camera. Touch the buttons on the left hand’s palm using your right hand’s index finger” at the top, and the camera feed in full screen. Once a left hand was detected, the system would display a menu on the palm with four items: View Map (of the local area), Food Menu (of a nearby restaurant), Bus Routes (with timetables of the university bus stop), and Weather Today (weather summary of the current week). Once an item is selected, the system displayed the content on the right side of the screen. I did not intervene or actively encouraged anyone to interact with PalmSpace. Observations were made from a distance, and I collected feedback if users actively looked around to discuss the concept they had just used and were available to provide user feedback. I ran this observational study for 2 hours (morning) in the university building’s main entrance and 2 hours (lunch time) in the restaurant.
Figure 8.1: In the wild demonstration in a (a) university building’s main hall, (b) in a restaurant, and (c) a user interacting with PalmSpace

8.2 Participants and Apparatus

The participants were anyone interested in trying out the demonstration. They mostly were university students. The apparatus uses the same server setup as previous studies. The web application is modified to display a menu of the left palm, which the users can use to view content on the right side of the web app. For example, the users can view the map of the nearby location, Food Menu of the nearby restaurant, weather summary of the week, and common local bus routes.
8.3 Procedure

A poster was set there to encourage interested people to interact with the system. A label at the top of the web app displayed instructions on how to initiate the menu with their left palm and use the right hand’s index finger to select the options. The instructions were in English. The users were observed from a distance. Feedback was also collected after the users stopped interacting with the demo application. This observational study was run for two days during lunchtime.

Figure 8.2: Menu options: (a) online map, (b) weather report rendered in real-time inside iframe HTML element. © Pinku Deb Nath, 2022.

8.4 Feedback and Observations

I received a total of 166 successful palm detection events, lasting around 8 seconds on average. Of these 166 palm detection events, 118 were item selections (≈ 0.7 selection / palm detection). From observing the statistics, I can conjecture that most users actually selected an item once the menu triggered. I observed that there were more engaged participants in
8.5 DISCUSSION

the restaurant than in the hallway. Several participants were holding beverages during both the sessions at the hall and the restaurant. The interested participants at the hall did not have any space for keeping their beverages to free up their hands, while the participants at the restaurant could easily place their drinks on a table and interact with the application. When in the hall, participants simply left as there was no solution to put down the beverage required for the two-handed interaction. This corroborated the conclusion that placing public interface displaying content in leisure-oriented environments, such as swimming halls, garner more attention than in business-oriented environments [5]. Hence, location is an important factor to consider when deploying public interfaces [6].

From the observations during interaction and informal discussions, I gathered that several participants successfully used PalmSpace by combining palm for menu appearance and dwell for item selection while facing the screen. Interestingly, several participants tried to perform a click action at first but quickly realized the presence of a timer at the bottom of the screen for the dwell action.

The overall system was engaging. Two commented that it looked “cool” and one thought that it “works really well”. Several participants joined in group, often starting by waving at the camera as they saw themselves in the live video feed. Users interested in the prototype often asked about the research project’s goals. They were impressed by the fact that the application was simply running on a standard laptop in a web browser. I received positive comments, such as “I am having fun!”. Lastly, the restaurant owner appreciated seeing so many customers engaged in this atypical activity in his establishment: “That is cool, there are many customers playing with your system!” The phenomenon was similar to the honeypot effect [93] in which the presence of engaged users increases the curiosity of the nearby people.

8.5 Discussion

Interestingly, users showed us a new way to use PalmSpace: Using the left hand to display the menu and the right hand to point in front of the palm, but toward the screen. This solution is less comfortable than actually touching the palm. I believe that, instead of textual instructions, other forms of guidance could have led to a better understanding of the pointing action (e.g., pictures or figures, or even animations).

Overall, users were keen to interact with PalmSpace and liked the concept. I think that reasons include curiosity and a trendy effect with an
augmented reality application. Nevertheless, the system was engaging users without a required or necessary task. This is encouraging, and I conjecture that PalmSpace can be used in real situations with actual needs (e.g., buying a transit ticket).
Chapter 9

Design Guidelines, Limitations and Future Work

I present my findings as design guidelines for palm interaction based on the study results.

9.1 Design Guidelines

9.1.1 Palm as Input Space

I suggest designers to consider the palm as an alternate input space as selecting items on the palm area is faster than in mid-air. Palm has a smaller area to access items than mid-air input. However, interactions on the palm should be short as prolonged usage of both hands in front of the device camera may trigger arm fatigue. Currently, designing interactions on the palm are limited by the robustness of the sensing technologies, which are continuously upgrading to include more accurate, versatile, and portable sensing technologies. Hence, the palm area will be promising in the near future for designing unique interaction techniques.

9.1.2 Location, Grid Size, and Layout

The palm area allows faster item selection than the finger area, possibly because the consistent area and coverage by the palm area allow the users to remember and recall target locations based on mentally assigned landmarks on the palm. In addition, with targets in a grid layout, no more than 3 items should be placed in a row to prevent significant performance degradation. However, there are situations where the actual on-screen has four or five items. I observed that users’ performance significantly degrades with more than 3 items in a row. In such instances, designers should consider rearranging items in multiple rows and can use either grid or flow for the layout.
9.1.3 Item Placement

When there is one group of items that need to be mapped onto the palm, we should prefer the palm area over finger for placing the items of the single group. If there are two groups stacked on top of one another then we should place the top group in the finger area and the bottom group in the palm area. Also, participants prefer mirroring the relative locations of the items within their groups especially when there is enough space for mapping all the items in the groups.

9.1.4 Presentation and Remediation

Item selection is faster when items are directly superimposed on the palm space. Thus, I suggest superimposing UIs directly on the palm via the camera feed for faster item selection over superimposing a virtual palm in UIs. When placing items, the participants prefer the left-to-right scanning direction. Privacy should be considered when superimposing UIs on the palm space and using the camera feed since few participants raised their concerns during the study.

9.1.5 Discoverability

Designers should consider providing instructions (e.g., animations or videos) to demonstrate how to initiate the interface with their palms. Once users initiate the interface, they can intuitively figure out how to use it.

9.2 Limitations

In this section, I will discuss some inherent limitations of PalmSpace and discuss possible future extensions of PalmSpace.

9.2.1 Prototype

The performance and reliability of PalmSpace depend on the quality of the hand-tracking system. Currently, PalmSpace uses conventional webcams with hand-detection modules provided by MediaPipe. The webcam quality and internet connection speed (for remote server access) can potentially influence the prototype’s performance. The variations in the resolution and the frame rate of the webcams could potentially affect the robustness and accuracy of the hands detection model. In addition, environmental factors could adversely affect hand detections. For example, we observed that
9.3. **FUTURE WORK**

Bright light facing directly on the webcam could saturate the pixels in the images and cause errors in detection. However, we found that the model can detect sufficiently well in low light conditions.

I used Google’s MediaPipe to track users’ finger positions and leverage the data to allow users to interact with PalmSpace. Though Mediapipe provides reliable 2D positions of the hand and finger locations, I found that hand/finger depths were unreliable as the images were commonly captured with built-in RGB cameras. Initially, I started implementing tap based on the relative information of the fingertips with respect to the location of the base of the palm of the same hand. However, the relative depth information in relationship with both hands was unreliable. High-resolution off-the-shelf depth sensors could have mitigated this issue. I anticipate that advancements in cameras will enable more reliable tracking to trigger reliable selections. Future work can explore new trigger gestures with enhanced prototypes.

9.2.2 Number of Participants

We recruited a limited number of participants for our user studies (e.g., 12 participants for each study except the initial survey). We acknowledge that larger participant pools from diverse backgrounds may better represent the variances in user performance and elucidate more diverse preferences and observations. In addition, it will allow us to analyse data based on gender, users’ palm size and other related factors.

9.3 Future Work

9.3.1 Guidance to novice users

I explored options using a screen to provide visual feedback, i.e., for novice users. With a robust hand-detection mechanism, future work can also explore the novice to expert transition and the relations between memorability, items location, and palm landmarks (e.g., joints). Additional work can be directed towards exploring whether such visual guidance can be superseded by memorizing target placements patterns using fixed landmarks on the palm. This task is dependent on the robustness of the hand-detection mechanism.
9.3. FUTURE WORK

9.3.2 Reducing Fatigue

In study 3, participants reported that they experienced fatigue while using PalmSpace. One approach to minimize fatigue is to reduce interaction duration. Another approach could be adjust cameras to track users hand in any natural positions that triggers less fatigue. Therefore, future work can explore alternative methods of palm interaction to reduce fatigue.

9.3.3 Expanding interaction space

The current prototype uses the left hand for anchoring the input space and the index finger of the right hand to control the cursor. Future work can explore how to include multiple cursors using multiple fingers to allow more expressive gestures like touch pads on laptops. I investigated 2, 4 and 6 items in a group for item placement. Future work can explore ways to accommodate more than 6 items with strategies such as using items in multiple layers.

From my knowledge, this research is among the first to explore the palm area as an input space using only a commercially available camera with no depth-sensing capabilities. I intentionally considered a few factors to make the prototype relevant for a wide range of public interfaces and conducted remote user studies. However, further research can explore other factors that will aid future endeavors for exploring the palm’s interaction space.

Future work can also explore multi-layer UIs in the palm and compare performance on different tasks with on-screen UIs. Researchers can examine the effects of iterating through multiple UIs by selecting items that force the items to have a small interaction surface.

Researchers can also explore adaptive UIs that adjust different UI elements based on the user’s interaction patterns. For instance, adaptive UIs can update the location and size of items based on how frequently the items are selected by the users or update items’ locations based on the users’ interaction styles.
Chapter 10

Conclusion

In this thesis, I took an initial step to explore PalmSpace, a viable input technique that leverages users' hand palm to interact with public touchscreen devices. I conducted a series of user studies for evaluating the factors necessary for designing PalmSpace. First, I conducted design studies to ascertain the preferences regarding different design factors for interfaces located in the palm area and expectations for alternatives to public touchscreen interfaces. The design studies included a survey and a pilot study. In the survey, I explored different factors associated with designing PalmSpace. I further ran a pilot study to determine the factors for item placements. Based on my survey and pilot results, I developed a prototype PalmSpace allowing users to interact with a device using their palm. The prototype consists of a web application that ran the MediaPipe hand-tracking module [87] in real-time and used video from conventional webcams. The application was designed in a self-contained manner so that there were no special hardware requirements apart from a laptop and internet connection - making the prototype suitable for low-cost implementation in public spaces.

I conducted three user studies to determine which factors impact target selection times and user preferences. I iteratively made modifications to the prototype based on the study requirements. The first two of the studies focused on finding appropriate parameters for design factors while the third study evaluated PalmSpace with mid-air interaction, and the fourth study evaluated the user acceptance of PalmSpace in public places. I investigated design factors and their parameters such as suitable remediation, grid size, and location in study 1. Results showed that direct mapping of interface onto the palm (S2H), up to 3 items per row/column, and palm for grid location to be the best choices. Study 2 focused on determining the parameters for layout when there are five items in a row, scanning direction for selecting the on-screen items for mapping and integration of the PalmSpace interface with the existing UI. Results revealed that choices for layout did not matter while left-to-right scanning direction and full integration were preferred for scanning direction and integration, respectively. In the third study, I compared PalmSpace with MidAir interaction for completing a sequence of tasks.
representing the scenario of booking train tickets and found PalmSpace to be faster than MidAir.

I also ran a demo in the wild - in two public spaces - to collect users' feedback. The PalmSpace prototype was set up in a restaurant and in a hallway, and permitted to view food menus, maps, train routes, and weather. People interacted with the system in the public space and provided me with helpful feedback on the PalmSpace and the application. I observed that the users need to be guided to get accustomed to the new PalmSpace interface. Once the users got accustomed, they stayed engaged, especially within a friendly crowd.

Finally, I discussed design guidelines for palm interaction, limitations of our current work, and future work ideas. Prior research for mid-air and on-body interaction used specialized and wearable sensors that are not possible to use for regular use in public interfaces due to many factors such as expensive setup and the need to wear additional devices. I limited my hardware requirements to what will be available to most users and depended more on software-based solutions for hand-tracking and providing feedback. I expect that sensing additional high-resolution information such as per-pixel depth information will be possible with the advances in cameras, wearable and sensing technologies, and their integration into commercial products. In addition, given the various health risks of interacting with public interfaces, there will be an incentive to develop different solutions that will help to mitigate and isolate risk factors. Hence, research for designing different interfaces such as PalmSpace that can use emergent advanced sensing technologies will be important in the near future. My thesis work will serve as a stepping stone for future researchers to accelerate the development of such interfaces.
Bibliography


[28] **Freeman, E., Brewster, S., and Lantz, V.** Tactile feedback for above-device gesture interfaces: Adding touch to touchless interactions. In *Proceedings of the 16th International Conference on Multi-
**Bibliography**

*modal Interaction* (New York, NY, USA, 2014), ICMI ’14, Association for Computing Machinery, p. 419426. → pages 11


72


Bibliography


76


Conference on Human-Computer Interaction with Mobile Devices and Services (New York, NY, USA, 2009), MobileHCI 09, Association for Computing Machinery. → pages 7, 12


[99] Niikura, T., Hirobe, Y., Cassinelli, A., Watanabe, Y., Komuro, T., and Ishikawa, M. In-air typing interface for mobile devices with vibration feedback. In ACM SIGGRAPH 2010 Emerging Technologies (New York, NY, USA, 2010), SIGGRAPH 10, Association for Computing Machinery. → pages 8, 11, 12


[113] Song, J., Sörös, G., Pece, F., Fanello, S. R., Izadi, S., Keskini, C., and Hilliges, O. In-air gestures around unmodified mobile


84


[137] Yeo, H.-S., Feng, W., and Huang, M. X. WATouCH: Enabling Direct Input on Non-Touchscreen Using Smartwatch’s Photoplethys-


Appendices

Appendix A

Design Exploration

A.1 Combinations of on-screen items

For design exploration pilot study, I considered several combinations of on-screen items arranged in one or two groups in different orientations. The combinations demonstrated the different combinations of the factors Number of items and Item arrangement. Figure A.1 and Figure A.2 display few of those possible combinations for two groups.

A.2 Demonstration

The following sections provide few examples summarizing different layout information for one and two groups. Each of the examples for the layouts contains an initial state, a single user input and aggregated results for those layouts.

A.2.1 One group

Figure A.3 demonstrates the combinations and results for the considered grid arrangements for one group with 6 items. Figure A.3 (a) shows the initial state of items in a 2-row grid layout and (d) shows the initial state of items in a 3-row grid layout, and both of them represent items in one group. Figure A.3 (b) and (e) show the initial states for the two grid layouts with 2 and 3 rows. Analogous to the previous figure, Figure A.3 (b) and (e) demonstrate samples of the preferences provided by one of the six participants while Figure 4.3 (c) and (f) show the aggregated preferences of all the participants for the respective combinations.

A.2.2 Two groups

Figure A.4 demonstrates the combination and result for 6 items arranged in grids of 2 rows within the two groups arranged vertically. Figure A.4a represents the initial state, Figure A.4b represents a sample of preference
A.2. Demonstration

Figure A.1: Combinations of on-screen items arranged in two groups with 4 items per group: (a) 4 items with horizontal layout per group and the groups arranged horizontally, (b) 4 items with horizontal layout per group and the groups arranged vertically, (c) 4 items with vertical layout per group and the groups arranged horizontally, (d) 4 items with vertical layout per group and the groups arranged vertically, (e) 4 items with a grid layout of 2 rows per group and the groups arranged horizontally, (f) 4 items with a grid layout of 2 rows per group and the groups arranged vertically.

provided by one of the participants while Figure A.4c represents the aggregated preferences of all 6 participants.
A.2. Demonstration

Figure A.2: Combinations of on-screen items arranged in 2 groups with 6 items per group: (a) 6 items with horizontal layout per group and the groups arranged horizontally, (b) 6 items with horizontal layout per group and the groups arranged vertically, (c) 6 items with vertical layout per group and the groups arranged horizontally, (d) 6 items with vertical layout per group and the groups arranged vertically, (e) 6 items with a grid layout of 2 rows per group and the groups arranged horizontally, (f) 6 items with a grid layout of 2 rows per group and the groups arranged vertically, (g) 6 items with a grid layout of 3 rows per group and the groups arranged horizontally, (h) 6 items with a grid layout of 3 rows per group and the groups arranged vertically.
A.2. Demonstration

Figure A.3: One Group, Grid Arrangement (a, b, c) and (d, e, f) represents two ways in which grid arrangements are considered for one group, (a, d) are initial states, (b, e) shows the result after a participant maps the items onto the palm image while (c, f) are the aggregated results of all the six participants’ input for the considered combinations.
A.2. Demonstration

Figure A.4: Two groups representing vertical group arrangement, 6 items arranged in a grid of 2 rows within the two groups, (a) are initial states, (b) shows the result after a participant maps the items onto the palm image while (c) are the aggregated results of all the six participants’ input for the considered combinations.