Gesture Design and Implementation for Interacting on Surfaces

François Bérard
Spring school of the 3ème cycle Romand in Computer Science
June 2010
What is it about?

Interacting with new devices with no keyboard and no mouse
What is it about?

Interacting with new devices with no keyboard and no mouse

- PDA
- Tablet PC
- Large interactive surfaces
What is it about?

Creating New Forms of General Purpose Human-Computer Interactions

- Interaction is explicit
- Interactions are generic
  one interaction can be used to achieve different tasks
- Tasks are productive
Why would you want to do that?

Not many other options

A keyboard and a mouse on my iPhone?
Why would you want to do that?

Because is it more *direct*


**Direct Manipulation Interfaces**

- Windows Icons Mouse Pointer (WIMP) interfaces

**Direct Touch**

- post-WIMP
Why would you want to do that?

Because is it more *natural*

- *Natural User Interface* is the new buzzword in HCI
- What is a Natural user interface?
- Is Natural *better* for user interaction? for example: more efficient?
The User Centered Approach

What is the benefit to users?

• Identify a problem / a need / a task
• Design a new solution
• Evaluate the added value

What it is not

• Technology driven
• Marketing / fashion / novelty driven
Overview

Standing on the shoulder of giants: past work

State of the Art

Implementation approaches
Standing on the Shoulders of Giants:
Literature Overview
Literature Overview

Foundations
Command activation
Two handed interaction
Tangible Interaction
Foundations

About Gestures


3 functions of gesture

• ergotic function: material action, modification and transformation of the environment.

• epistemic function: perception of the environment.

• semiotic function: communication of information towards the environment.
Foundations

About Gestures


3 functions of gesture in WIMP interfaces

• ergotic: moving a file to a folder, activating a button

• epistemic function: finding the links in a (badly designed) web page

• semiotic function: activating a command
Foundations

The Fitts’ Law


\[ \text{Mean Acquisition Time } MT = a + b \log_2(D/W + 1) \]

\[ \text{Mean Acquisition Time } MT = a + \frac{1}{IP} \cdot ID \]
Foundations

The Fitts’ Law

\[ MT = a + \frac{1}{IP} \cdot ID \]

[MacKenzie92]
Command Activation

Pie Menus


- Linear Menus
  - Acquisition of 2D targets with high
    Index of Difficulty
  - Difficulty varies between options

- Pie Menus
  - Better compatibility (angle, time, opposed commands)
  - Constant distance to option
  - Selection distance smaller
  - But reduced nb. of options ➔ hierarchical menus
Command Activation

Marking Menus

Command Activation

Marking Menus

Command Activation

Marking Menus

Command Activation

Control Menus


Many commands require parameters

- Font size
- Zoom factor

Menus (pop-up / pie / marking) activate only, require additional interactions for the parameters (dialog box).

Control Menus combine activation and parameters in a single gesture
Command Activation

Control Menus

Putting it all together

BumpTop

Text Input

Semiotic vs. Ergotic: Graffiti vs. Virtual Keyboard

Unistroke

Text Input

Semiotic vs. Ergotic: Graffiti vs. Virtual Keyboard

Graffiti

Text Input

Semiotic vs. Ergotic: Graffiti vs. Virtual Keyboard

Exploiting Human Dexterity

Motivations

WIMP (Windows Icons Mouse Pointer)
Input restricted to the mouse (2D, binary buttons)
and keyboard (binary buttons)

Physical World
Dual-hand interactions
Complex handling with all fingers (high degrees of liberty)
Two Handed Interaction


We use both hands to interact with objects in the real world, why not with virtual objects?
Two Handed Interaction

Single hand vs. two-handed interaction

Two Handed Interaction

Single hand vs. two-handed interaction


Hypothesis

Two-Handed Interaction is more efficient because of:
• Division of labor: each hand remains in the neighborhood of its task
• Parallelism
## Two Handed Interaction

**Single hand vs. two-handed interaction**

<table>
<thead>
<tr>
<th></th>
<th>Total (s.)</th>
<th>Use of non-Dominant</th>
</tr>
</thead>
<tbody>
<tr>
<td>Floating - Dominant</td>
<td>2.89</td>
<td>--</td>
</tr>
<tr>
<td>Floating - non-dominant</td>
<td>2.96</td>
<td>49%</td>
</tr>
<tr>
<td>Palette</td>
<td>2.90</td>
<td>47%</td>
</tr>
<tr>
<td>Toolglass</td>
<td>2.43</td>
<td>83%</td>
</tr>
</tbody>
</table>
Two Handed Interaction

The Kinematic Chain Model


Study of bi-manual actions in the real world.

- Writing on paper
- Painting with a color palette
- Hammer a nail
- Threading a needle
Two Handed Interaction

The Kinematic Chain Model


Study of bi-manual actions in the real world.

- Writing on paper
Two Handed Interaction

The Kinematic Chain Model


- The non dominant hand defines the spatial reference for the dominant hand
- The spatial and temporal scale of motion is coarser for the non dominant hand
- The order of action is: non dominant, then dominant
Tangible Interaction


Using the epistemic function of gestures

Direct Manipulation Interfaces

Direct Touch

Tangible Interaction
Tangible Interaction


Using the epistemic function of gestures
Tangible Interaction

Superiority of Passive Haptic Feedback vs Direct Touch
Putting it all together

The Magic Table


- Computers are ill-suited for informal meetings
- (unequipped) Whiteboard are still widespread
Putting it all together

The Magic Table

Removing all Acquisition Steps

Direct Manipulation Interfaces

Direct Touch

Tangible Interaction

Perceptual Window
Removing all Acquisition Steps

The Perceptual Window

Removing all Acquisition Steps

The Perceptual Window

State of the Art
Extending Marking Menus to Multi-Touch

Improving the Learnability of Gestures

Octopocus

Improving the Learnability of Gestures

Shadowguides

Finding a Set of Intuitive Gestures

User Defined Gestures


How can we find the best gesture set on large interactive surfaces?

1. Show a referent (and speak)
2. Ask for a sign
Finding a Set of Intuitive Gestures

User Defined Gestures

- Experts elicit a set of 27 referents

<table>
<thead>
<tr>
<th></th>
<th>Gesture</th>
<th></th>
<th>Gesture</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Move a little</td>
<td>15</td>
<td>Previous</td>
</tr>
<tr>
<td>2</td>
<td>Move a lot</td>
<td>16</td>
<td>Next</td>
</tr>
<tr>
<td>3</td>
<td>Select single</td>
<td>17</td>
<td>Insert</td>
</tr>
<tr>
<td>4</td>
<td>Rotate</td>
<td>18</td>
<td>Maximize</td>
</tr>
<tr>
<td>5</td>
<td>Shrink</td>
<td>19</td>
<td>Paste</td>
</tr>
<tr>
<td>6</td>
<td>Delete</td>
<td>20</td>
<td>Minimize</td>
</tr>
<tr>
<td>7</td>
<td>Enlarge</td>
<td>21</td>
<td>Cut</td>
</tr>
<tr>
<td>8</td>
<td>Pan</td>
<td>22</td>
<td>Accept</td>
</tr>
<tr>
<td>9</td>
<td>Close</td>
<td>23</td>
<td>Reject</td>
</tr>
<tr>
<td>10</td>
<td>Zoom in</td>
<td>24</td>
<td>Menu access</td>
</tr>
<tr>
<td>11</td>
<td>Zoom out</td>
<td>25</td>
<td>Help</td>
</tr>
<tr>
<td>12</td>
<td>Select group</td>
<td>26</td>
<td>Task switch</td>
</tr>
<tr>
<td>13</td>
<td>Open</td>
<td>27</td>
<td>Undo</td>
</tr>
<tr>
<td>14</td>
<td>Duplicate</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Finding a Set of Intuitive Gestures

User Defined Gestures

- Experts elicit a set of 27 referents
- Participants are presented the referents and asked to perform gestures (one handed, dual handed)

- 20 paid participants (9 females, 11 males)
- General public (no computer scientist, designers)
- No prior experience of tabletop, iPhone, etc. (dual touch)
- All right handed
Finding a Set of Intuitive Gestures

User Defined Gestures

- Experts elicit a set of 27 referents
- Participants are presented the referents and asked to perform gestures (one handed, dual handed)
- Performed Gestures are grouped, agreement scores are computed
- Gestures with best agreement are retained
Finding a Set of Intuitive Gestures

User Defined Gestures
Finding a Set of Intuitive Gestures

User Defined Gestures
Finding a Set of Intuitive Gestures

User Defined Gestures

- Participant prefer single finger gestures.
- Some referents are too complex for a gesture: need for widgets?
- Gestures with strong agreement had higher subjective rating.
Don Norman’s Viewpoint

On Natural User Interfaces


- Graphical User Interfaces made the interface visible, discoverable
- Most gestures are neither natural nor easy to learn or remember
- All new technologies have their proper place
- “Are natural user interfaces natural? No. But they will be useful.”
Implementation Approaches

Tracking Fingers
Capacitive Sensing

SmartSkin

Capacitive Sensing

MERL DiamondTouch

Capacitive Sensing

MERL DiamondTouch

Capacitive Sensing

DiamondTouch Single User Multi-Touch

Capacitive Sensing

DiamondTouch Single User Multi-Touch
Capacitive Sensing

DiamondTouch Single User Multi-Touch
Capacitive Sensing

DiamondTouch Single User Multi-Touch

- Track 1D peaks

![Graph showing track of peaks](image)

- Predict new position from last position and speed

\[ \hat{p}_t = p_{t-1} + s_{t-1} \times \Delta t \]

- Associate observations to closest prediction
Capacitive Sensing

DiamondTouch Single User Multi-Touch
Capacitive Sensing

DiamondTouch Single User Multi-Touch

- Modeling peaks as a mixture of gaussians
Capacitive Sensing

DiamondTouch Single User Multi-Touch

- Modeling peaks as a mixture of gaussians
  - Better Accuracy
Capacitive Sensing

DiamondTouch Single User Multi-Touch

- Modeling peaks as a mixture of gaussians
  - Better Accuracy
  - Better handling of shadowing
Capacitive Sensing

DiamondTouch Single User Multi-Touch

- Prediction update
- Thresholded peaks computation
- Thresholded peaks association
- Mixture of Gaussians computation
- Position and Speed computation

\[ \hat{p}_t = p_{t-1} + s_{t-1} \times \Delta t \]
Capacitive Sensing

DiamondTouch Single User Multi-Touch

- 2D Tracker
  Associates 2 x 1D tracked peaks into a 2D touch
Capacitive Sensing

DiamondTouch Single User Multi-Touch

- 2D Tracker
  Associates 2 x 1D tracked peaks into a 2D touch
Capacitive Sensing
Computer Vision

Direct Vision

Computer Vision

Direct Vision


4 stages
• Foreground / background computation
Computer Vision

Direct Vision


4 stages

• Foreground / background computation
• Thresholding
Computer Vision

Direct Vision


4 stages
- Foreground / background computation
- Thresholding
- Connected components analysis
Computer Vision

Direct Vision


4 stages

• Foreground / background computation
• Thresholding
• Connected components analysis
• Shape Filtering
Computer Vision

Direct Vision


4 stages

- Foreground / background computation
- Thresholding
- Connected components analysis
- Shape Filtering
- Association (tracking)
Computer Vision

Direct Vision

• Foreground / background computation

\[ p = [R, G, B] \]
\[ [r, g] = \left[ \frac{R}{R + G + B}, \frac{G}{R + G + B} \right] \]
\[ d(p, p') = ||[r, g] - [r', g']|| \]
Computer Vision

Direct Vision

- Thresholding

\[
m_0 = \text{median}_{(x,y)} d^t(x,y)
\]
\[
m_1 = \text{median}_{(x,y)} \left| m_0 - d^t(x,y) \right|
\]
\[
\theta^t = m_0 + 4 \cdot m_1
\]
Computer Vision

Direct Vision

- Thresholding


$$\sigma_{\text{Within}}^2(T) = n_B(T) \cdot \sigma_B^2(T) + n_A(T) \cdot \sigma_A^2(T)$$
Computer Vision

Direct Vision

- Connected Component Analysis
Computer Vision

Direct Vision

- Shape Filtering
  - \( p \) is in the foreground
  - \( p \) is in a connected component of size \( \geq a \)
  - \( p \) is surrounded by a (filled) disk of diameter \( d \)
  - there is only one connected component on the perimeter \( C \)
  - the AB distance is compatible with the width of a finger
Computer Vision

Direct Vision

- Association

\[ \dot{P}_t = P_{t-1} \]

\[ \dot{P}_t = P_{t-1} + v_{t-1} \cdot \Delta_t \]
Computer Vision

Direct Vision

Computer Vision

Direct Vision

Computer Vision

InfraRed Vision: Direct Illumination from the Back

**Computer Vision**

InfraRed Vision: Direct Illumination from the Back

Computer Vision

InfraRed Vision: Frustrated Total Internal Reflection

Computer Vision

InfraRed Vision: Frustrated Total Internal Reflection

On Gestures for General Purpose Human-Computer Interaction

- Gestures make interaction more direct (especially when tangible)
- Gestures can improve the ease of learning and efficiency, if well designed
- Extending the expressivity of gestures is tricky
  Is it even relevant?