A Gaze Interactive Textual Smartwatch Interface

John Paulin Hansen
Janus Aksø Madsen
Morten Jonassen
Haakon Lund
Javier San Agustin
Sebastian Sztuk
Speech and gaze goes well together
Equipment*

30 Hz Eye Tribe USB 3 Tracker  
Accuracy 1.4 Precision 1.0 - stationary

Track box: X = 30 cm, Y = 20 cm and Z = 12 cm at 35 cm distance

A 1.8” Adafruit monochrome TFT screen with a resolution of 160 x 128 pixels, controlled by an Arduino Nano board

The Eye Tribes Java Client with a custom mobile gaze tracking framework

A standard PC

Velcro strips
SW tracking from glasses vs. a user facing camera in the watch

Figure from Akkil et al: CHI 2015.
Smooth pursuit

Augusto Esteves et al. UIST 2015
The textual interface

now.
Hands-free

- What if people had a gaze tracking unit with them all the time that they could easily use to control the environment?
4 virtual input fields
8 virtual input fields
Accuracy and precision – standing and walking slowly

<table>
<thead>
<tr>
<th>Condition</th>
<th>Accuracy*</th>
<th>Precision*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Standing</td>
<td>2.93 ± 0.11</td>
<td>2.95 ± 0.09</td>
</tr>
<tr>
<td>Walking</td>
<td>3.00 ± 0.12</td>
<td>3.02 ± 0.10</td>
</tr>
</tbody>
</table>

Illustration – not real data
Gaze gestures are unnatural—but they work
10 people from our lab (6M, 4F)

Wearing SMI Eye Tracking Glasses 2

Twelve use-cases; e.g.:

“How would you move your eyes to decrease the temperature?”

“How would you move your eyes to unlock the door?”

“How would you move your eyes to go back in the menu on your smartwatch?”
Observations

- 86% of all enacted control actions would be done by (combinations of) going up, down, left or right
- The object controlled would always be attended during the imagined input sequence
  - 40% of the times as the first fixation
  - 60% during the sequence
Gaze interactive areas

- Speed up
- Speed down
- Back
- Display area
- RSVP
- PICTURE >>
Gaze gestures

- Speed up
- Speed down
- Back
- RSVP
- Display area
Look-away command
Observations

• 27 subjects (mean age 29 years, 14 women)
• First they did a training task with a tablet pen then with gaze
• It took 231.4 seconds (S.D. = 57.1 s) to complete the task with gaze.
• Hit rates:
  – jumping back one sentence 70 % (S.D. = 45%);
  – pictures were 100 % correctly activated 100 % (S.D. = 0 %).
• The RSVP speed levels were adjusted to an average of 174 words per minute.
• 9 preferred the pen input but a majority of 18 participants preferred gaze input, mostly because they said:
  – it was fast and easy,**
  – it made the text stop streaming when not a9 ended,*
  – it felt less strenuous than holding a stylus pen up to the tablet* display *
LED leeding lights
GazeWatch interaction with public displays