Faking magic

Wizard of Oz experiment

- Unseen human plays the part of the not-yet-extant computer system, allowing dynamic interaction features to be studied.
  - The Wizard of Oz operates fake apparatus from behind a curtain.
- "Experimenter-in-the-Loop" paradigm.
- Behaviour of A.I. system simulated by a human (opposite of usual goal).

Origin

- J. F. Kelley coined the terms "Wizard of Oz" and "OZ Paradigm" when he was working on his dissertation in the early 1980s.

Understanding user performance in practice

• Typically still under controlled conditions (e.g. laboratory).
• Initial analysis may take the form of semistructured pilots, during which participants perform a task with simulated system support.
• Faster and cheaper than developing a fully functional system.
• Must still remain believable and appear realistic though!


Knowledge-based process for multimodal interface design.

• No “real” system before the Design Rationale cycle is complete.

(Web GUI and voice design


1. BM: initial Behavioural Model consisting of information acquired from existing graphical and voice interface designs and the inclusion of new requirements as the identification of new needs.
2. TA: perform Task Analysis to get the representation of input and output information and to apply an appropriate method.
3. TR: abstract Task Representations of the necessary tasks information in the interface context enabling a comparison between solutions.
4. AS: Analyse the Solutions incorporating any Design Rationale captured earlier and filtering out the interaction options. This step provides the potential modalities for the user to perform the task.
5. WOz : Wizard of Oz generation that will be evaluated with real users. The aim is to obtain a list of requirements and directives for the usability test that will be performed on the WOz interface.

First JAST P.o.C. study

• Before anyone went near a computer, never mind a robot.
2D photo stimuli

4 cameras and 4 mics

Mock version

- Experimenter had to act as “the system” controlling the environment, recording data (time, camera, audio, penalties), referee the rules.
- Allowed initial parameters, levels of difficulty, choice of modalities, length of experiment, etc. to be based on empirical evidence rather than entirely guess work.
- Generated rules to “force” joint collaborative action.
- Minimised the risk of buying expensive equipment and wasting many months of programmer time building a complex system that wasn’t very useful.

Next stage
**SpaceBook**

- Before artificial intelligence, artificial agents, ASR or anything like that was introduced...
- Started with human-human dialogue and moved towards the final scenario step-by-step.
- Start simple and build up.

**WOz**

- “Wizard of Oz” paradigm adopted to capture authentic dialogue and interaction data for system development purposes.
- Involved having a person take on the role of a SpaceBook User in the real (streets of Edinburgh) environment and interacting with a human who creates the illusion of being the SpaceBook system.
- Objective is to mimic the final SpaceBook procedure as closely as possible, only with a human agent playing the role of the artificial agent.
- Thus hypothetical scenarios can be investigated, issues of user design addressed and feasibility studies conducted, feeding the developmental process before a fully-functional system has to be completed.

**In particular**

1. to identify what sorts of tasks a pedestrian might undertake; to identify the types of information needed to support the execution of those various task.
   - Analyse real dialogues for use in computational modelling.
2. to assess how much the granularity of position affects that dialogue; to assess the quality of interaction given various ambient noise.
   - Find possible sources of error early.
3. Unusually perhaps, the Wizard is also a participant and changes for each dyad (different pairs for each test run).

**Three WOz phases**

1. Wizard follows (shadows) Tourist in the street.
   1. Shared real environment.
   2. Real-time two-way voice interaction.
2. Wizard remotely interacting (cyber-stalking) by phone call.
   1. Glympse app used to track GPS location.
   2. Real-time two-way voice interaction.
3. Wizard remote interacting by TTS.
   1. Google map location with streetview in browser.
   2. Custom TTS gui to send speech to phone.
Phase One

- Wizard has direct visual contact walking behind (approx 15 metres) but not direct communication.
- Shared the same visibility and environmental conditions.
- Tourist instructed to ignore the physical presence of the Wizard.
- Tourist had Android mobile phone and Bluetooth headset, operating hands- and eyes-free.
- Phone ran a GPS logging app (My Tracks).
- Standard phone call to Skype number.
- Wizard had wireless headset connected to a laptop (with built-in SIM slot) running Skype.

Tourist and System role-play

- Set a general task list with vague requirements to act as a framework (semi-structured).
- Run multiple instances with different people acting both roles.
- Analyse and look for common features as well as the wider range of possible things that may be said or done.

Background briefing

Similar to having a friend with excellent knowledge of the local area, the idea is that you are a tourist to Edinburgh. It is likely you have drawn on various sources of information to help you to decide what you are going to do whilst in Edinburgh. But your friend is a very useful source. Although their advice comes to you remotely, they have approximate knowledge of where you are. To provide a context, we have suggested some scenarios for you to ‘act out’. You should act out each scenario as you see fit. You may have many questions to ask, or you may have none. There is no limit to the breadth of questions you may ask – but your guide may not know everything. To make this as natural as possible, we have not suggested any specific questions you might ask. In many respects, you have to act out the scenarios, trying to imagine yourself as a tourist.

Tourist task list

1. You have registered early for a conference talking place in the Informatics Forum and now have most of a day to explore the nearby area before the late afternoon introductory session and evening reception takes place. So you’d like to start by visiting the National Museum of Scotland. That’s if it is open, of course.
2. You have just left the National Museum of Scotland and feel like a coffee. Make your way to somewhere that you can sit down and have a coffee.
3. You’ve heard a story about a statue of a dog that you think is nearby and would like to take a photo of the dog and perhaps learn a little more about the story.
4. You have arranged to have lunch with a friend in a nearby pub. You can’t remember the exact name but you are sure it had the word “Bell” in the title.
5. You arrive 10 minutes early, but realise that you probably don’t have enough cash. You do have your bank card with you, however.
6. After lunch you’d like to have a stroll in the nearest park or similar area. However, you’d also like to locate a (small) supermarket on the way.
7. Time for business. Head back to Informatics!
Data generated

• Telephone conversation recorded as time-stamped 128 bit MP3 files.
  • Stereo: one channel for Tourist; one channel for Wizard
  • Synchronised conversation but can be separated into left and right for transcription process.
• GPS data available in multiple formats.

<table>
<thead>
<tr>
<th>TIME</th>
<th>PATH</th>
<th>LATITUDE</th>
<th>LONGITUDE</th>
</tr>
</thead>
<tbody>
<tr>
<td>2016-09-29</td>
<td>123.456</td>
<td>123.456</td>
<td>123.456</td>
</tr>
</tbody>
</table>

Example of path recorded

Phase Two

• Similar equipment to Phase One only the laptop is now in a laboratory with ethernet connection and full-sized monitor.
• Remote monitoring using off-the-shelf apps:
  • Live location map (Glympse.com).
  • Browser window with Google Map and Street View which could be manually interacted with.
  • Screen capture software (Camtasia); video camera as back-up.

Wizard control
Example transcription segment

<table>
<thead>
<tr>
<th>Transcription Segment</th>
</tr>
</thead>
<tbody>
<tr>
<td>YouTube: User: What's the weather like? Is it going to rain?</td>
</tr>
<tr>
<td>User: Yes, it's going to rain today.</td>
</tr>
<tr>
<td>52°F, 89% chance of rain.</td>
</tr>
<tr>
<td>User: Alright. Thanks for the info.</td>
</tr>
</tbody>
</table>

Video
Phase Three

• Wizard now has a custom-built interface (WOz Tool) over twin screens.
• Wizard no longer spoke directly but used a text-to-speech messaging system.
  • System sound authenticity (synthetic voice).
  • Validate TTS component works.
  • Modify user behaviour – reduce tendency to “chatter” and engage in common phatic interaction which interferes with navigational and informational communication.
• Custom SpaceBook mobile phone app developed (VOIP mobile data not phone call). Audio + sensor data. [Real component.]
• Custom SpaceBook server [testing real component].

Text responses

1. Predefined fixed text strings, such as “Continue straight ahead”, that were activated via “hot buttons”.
2. Modifiable text strings that could be completed using predefined choice lists, as in Turn [choose: right, left].
3. Freely typed text strings, used when the former two were insufficient.
  • These text strings were then converted to speech on the pedestrians phone (using standard TTS).
Worked in Stockholm too!

WOz3 transcript sample

<table>
<thead>
<tr>
<th>#</th>
<th>Speaker</th>
<th>Line</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Speaker 1</td>
<td>I just keep walking in the direction you were going</td>
<td>0:00:00</td>
</tr>
<tr>
<td>2</td>
<td>Speaker 2</td>
<td>You make a turn at the corner of the street.</td>
<td>0:00:08</td>
</tr>
<tr>
<td>3</td>
<td>Speaker 1</td>
<td>Yes. How about me!</td>
<td>0:00:16</td>
</tr>
<tr>
<td>4</td>
<td>Speaker 2</td>
<td>That's great.</td>
<td>0:00:24</td>
</tr>
<tr>
<td>5</td>
<td>Speaker 1</td>
<td>The museum is the big building on your right.</td>
<td>0:00:32</td>
</tr>
<tr>
<td>6</td>
<td>Speaker 2</td>
<td>Yes, I'll fill me more about this museum. But it's opening hours,</td>
<td>0:00:40</td>
</tr>
<tr>
<td>7</td>
<td>Speaker 1</td>
<td>The entrance is just around the corner.</td>
<td>0:00:48</td>
</tr>
<tr>
<td>8</td>
<td>Speaker 2</td>
<td>What is it?</td>
<td>0:00:56</td>
</tr>
<tr>
<td>9</td>
<td>Speaker 1</td>
<td>Another museum in this area.</td>
<td>0:01:04</td>
</tr>
<tr>
<td>10</td>
<td>Speaker 2</td>
<td>In the direction you are going.</td>
<td>0:01:12</td>
</tr>
<tr>
<td>11</td>
<td>Speaker 1</td>
<td>I'll let you know.</td>
<td>0:01:20</td>
</tr>
<tr>
<td>12</td>
<td>Speaker 2</td>
<td>It's very interesting. Why not I walk around the corner.</td>
<td>0:01:28</td>
</tr>
<tr>
<td>13</td>
<td>Speaker 1</td>
<td>Another museum in this area.</td>
<td>0:01:36</td>
</tr>
<tr>
<td>14</td>
<td>Speaker 2</td>
<td>We can also see some art.</td>
<td>0:01:44</td>
</tr>
</tbody>
</table>

Combined XML data output
It's a great system,
but on windy days, I'm often blown off course...