case studies in design informatics

Lecture 4: Wizard of Oz testing

design informatics

NRlabs neuropolitics research

Robin Hill

Institute for Language, Cognition & Computation,
School of Informatics
and
Neuropolitics Research Lab,
Politics and I.R.
www.robin.org.uk
r.l.hill@ed.ac.uk

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

- Story: *The Wonderful Wizard of Oz* by L. Frank Baum. Book 1900; film 1939.
- J. F. Kelley coined the terms "Wizard of Oz" and "OZ Paradigm" when he was working on his dissertation in the early 1980s.

Kelley, J. F. (1984). An iterative design methodology for user-friendly natural-language office information applications. ACM Transactions on Office Information Systems, 2:1, 26-41.

[http://www.musicman.net/oz.html]

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!

Barthelmess, Paulo, & Oviatt, Sharon (2008). Chapter 12 - Multimodal Interfaces: Combining Interfaces to Accomplish a Single Task A2 - Kortum, Philip *HCI Beyond the GUI (pp. 391-444). San Francisco: Morgan Kaufmann.*

Web GUI and voice design

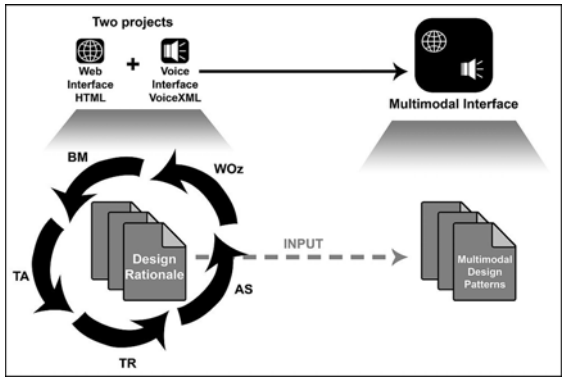
Neto, A. T., Fortes, R. P. M., & da Silva, A. G. (2008). Multimodal interfaces design issues: the fusion of well-designed voice and graphical user interfaces. In C. J. Costa, A. Protopsaltis, M. Aparicio & H. Oneill (Eds.), *Sigdoc'08: Proceedings of the 26th ACM International Conference on Design of Communication (pp. 277-278). New York: Assoc Computing Machinery.* DOI: 10.1145/1456536.1456597

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.

Knowledge-based process for multimodal interface design.

- No "real" system before the Design Rationale cycle is complete.

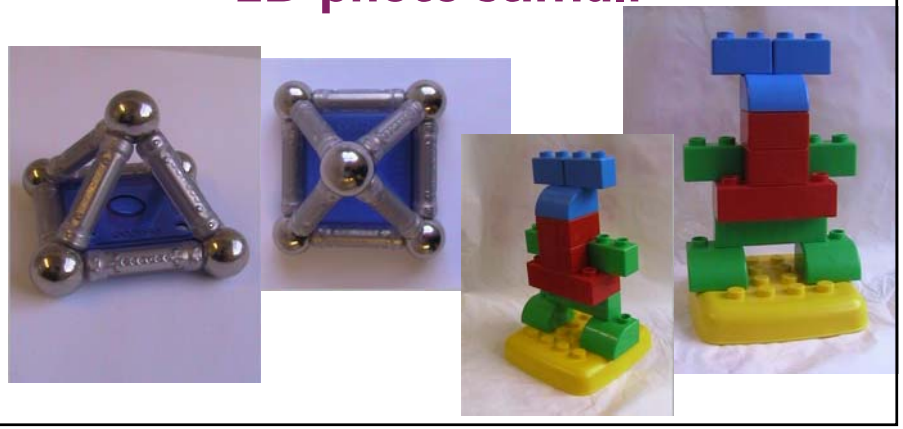
(Neto et al., 2008)



First JAST P.o.C. study

- Before anyone went near a computer, never mind a robot.

2D photo stimuli

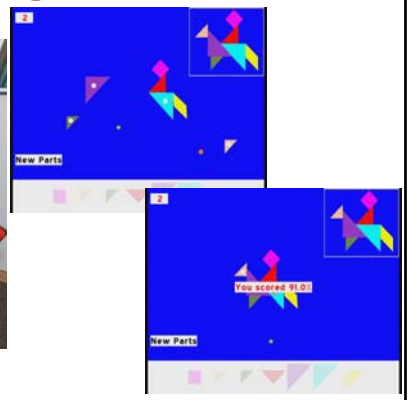


4 cameras and 4 mics

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.

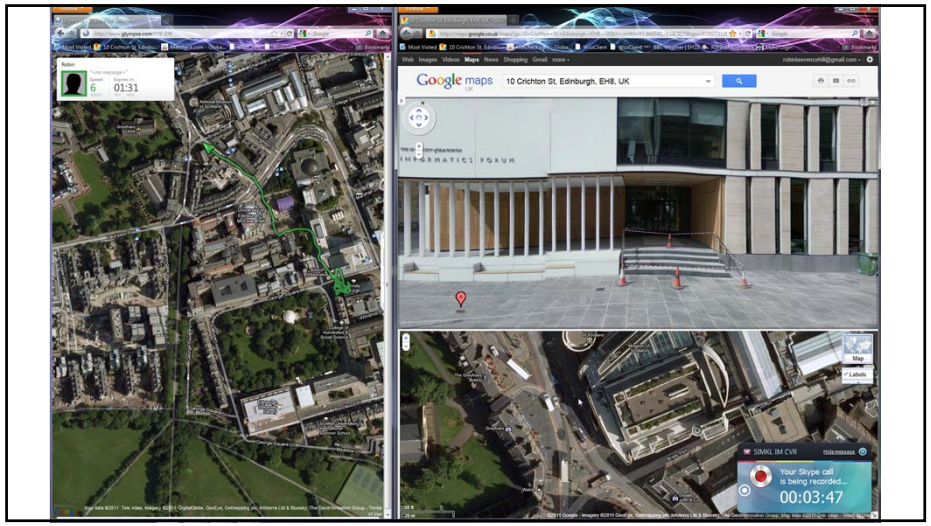
TYPE	TIME	LAT	LON	ALT	BEARING	ACCURAC	SPEED	NAME	DESCRIPTION	SEGMENT
TRACK								2011-07-15 11:14 Amy And Mary	Dyad 2	
P	2011-07-15T10:14:41Z	55.94461	-3.18755	160.3	0	20	0			0
P	2011-07-15T10:15:56Z	55.94459	-3.18753	160.1	0	12	0			0
P	2011-07-15T10:16:14Z	55.94452	-3.1876	165	184.7	10	0.5			0
P	2011-07-15T10:16:17Z	55.94448	-3.18773	165.6	184.7	9	0.75			0
P	2011-07-15T10:16:20Z	55.94453	-3.1877	168.5	184.7	9	0.25			0
P	2011-07-15T10:16:23Z	55.94456	-3.18772	169	184.7	13	0.5			0
P	2011-07-15T10:16:26Z	55.94464	-3.18777	168.5	349.4	11	1.25			0
P	2011-07-15T10:16:29Z	55.94471	-3.18781	168.2	348	10	1.25			0
P	2011-07-15T10:16:32Z	55.94476	-3.18783	169.2	348.2	13	1.25			0
P	2011-07-15T10:16:35Z	55.94484	-3.18787	169.1	346.5	12	1.25			0

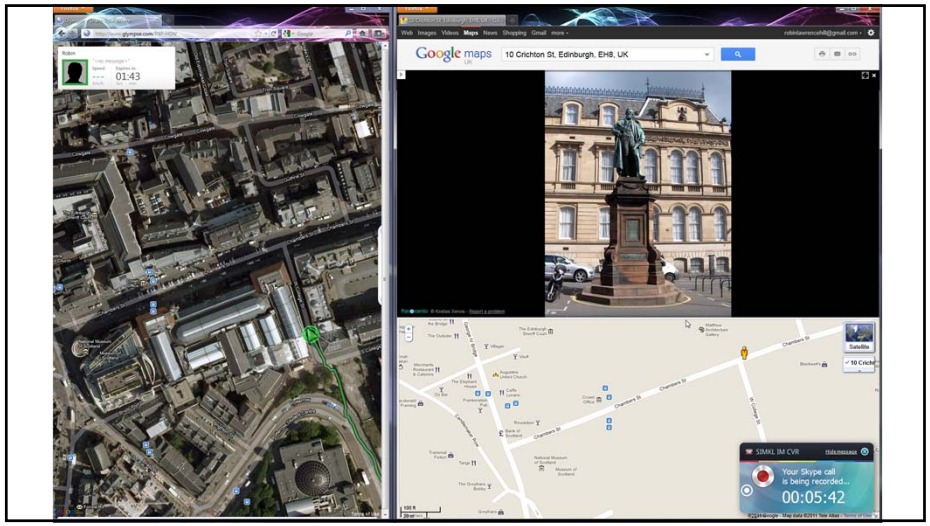
Example of path recorded

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

tourist		Um.
TC		00:00:50.330 - 00:00:51.630
wizard	Okay, well, straight ahead of you you'll see Bristo Square, where many of the university buildings are situated.	
TC	00:00:44.690 - 00:00:53.095	

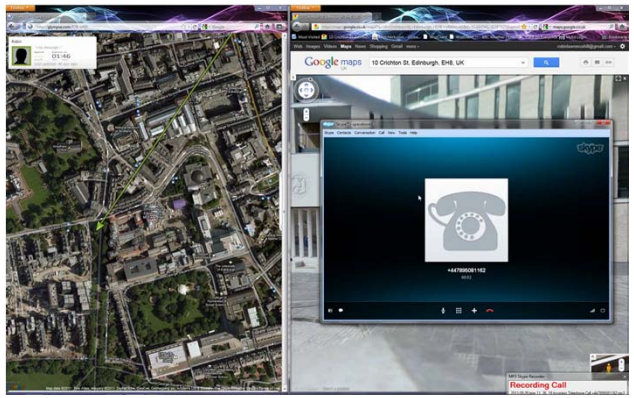
tourist	Uh, I'm sorry, your voice is breaking, could you please repeat it.
TC	00:00:53.070 - 00:00:56.600

wizard	Em, straight ahead of you you'll see Br- Bristo Square. If you continue walking straight ahead, you'll reach the square.
TC	00:00:56.220 - 00:01:03.480

tourist	Okay.	Okay, yeah, I'm g- I'm going forward. Uh. I see some um-. I think I'm seeing the Teviot place, the debating hall.
TC	00:01:00.733 - 00:01:03.190	00:01:03.445 - 00:01:11.100

tourist	And, uh, yeah. Should I move forward.		Okay okay I'm going there.
TC	00:01:11.214 - 00:01:14.099		00:01:16.504 - 00:01:18.700
wizard		Good yes, you're going in the right direction. (??)	
TC		00:01:14.120 - 00:01:17.330	

Video



Phase Three

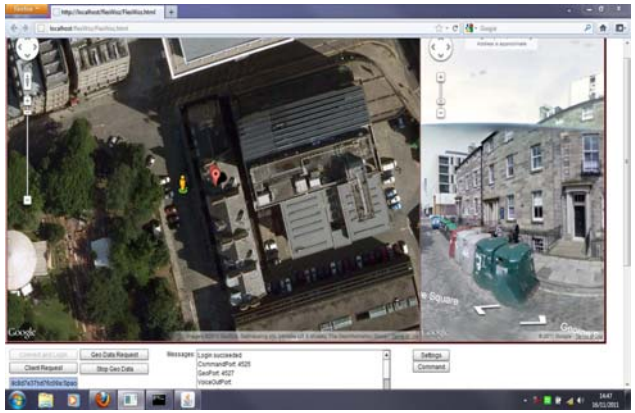
- Wizard now has a custom-built interface (WOz Tool) over twin screens.
- Wizard no longer spoke directly but used a text-to-speech messaging system.
  - System sound authenticity (synthetic voice).
  - Validate TTS component works.
  - Modify user behaviour – reduce tendency to "chatter" and engage 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

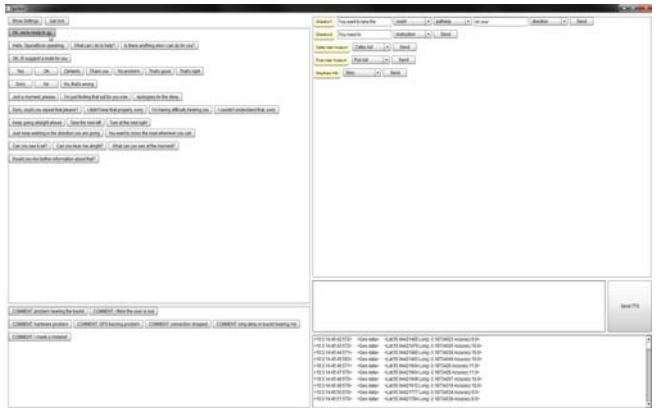
1. Predefined fixed text strings, such as "Continue straight ahead", that were activated via "hot buttons".
2. Modifiable text strings that could be completed using predefined choice lists, as in Turn [choose: right, left].
3. Freely typed text strings, used when the former two were insufficient.

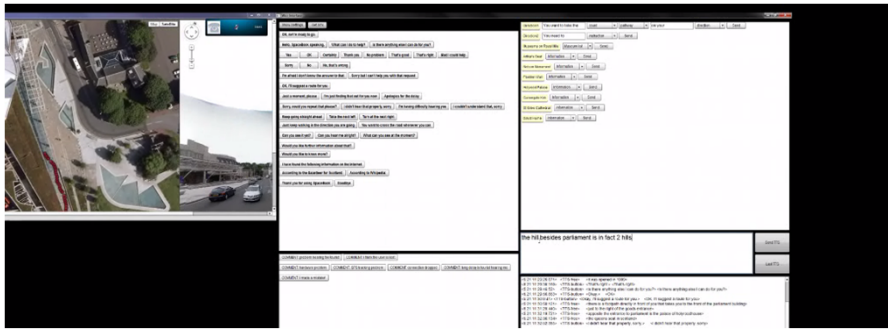
- These text strings were then converted to speech on the pedestrians phone (using standard TTS).

WOz Window 1

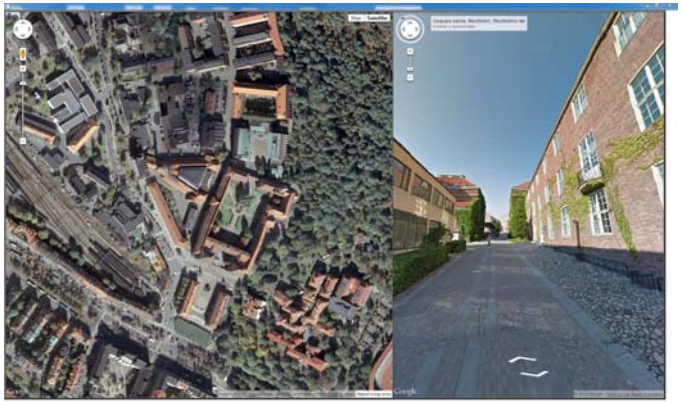


WOz Window 2





Worked in Stockholm too!

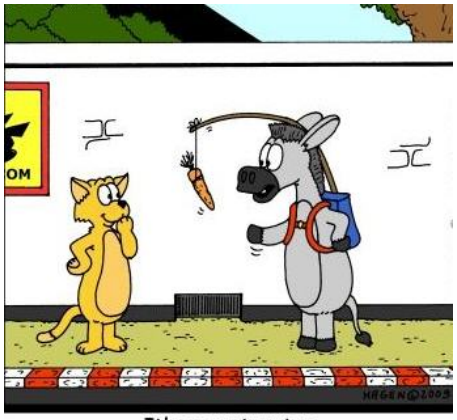


WOz3 transcript sample

275	SpaceBook	<TTS-button>	Just keep walking in the direction you are going
311	SpaceBook	<TTS-free>	You want the right hand side of the road
319	SpaceBook	<TTS-button>	Can you hear me alright?
321	tourist		Yep.
325	SpaceBook	<TTS-button>	That's good
353	SpaceBook	<TTS-free>	The museum is the large building on your right
355	tourist		Ah (??).
359	tourist		Can you tell me more about this museum, like its opening hours,
363	SpaceBook	<TTS-free>	The entrance is just around the corner
363	tourist		what there's to see inside,
365	tourist		do I need to pay an entrance fee?
377	SpaceBook	<TTS-free>	The museum opens at 10 am
381	SpaceBook	<TTS-free>	and closes at 5
404	SpaceBook	<TTS-free>	Is there anything else you would like to know?
407	tourist		Is there any, like, special exhibits going on in the museum currently?
426	SpaceBook	<TTS-free>	There is a special exhibition
434	SpaceBook	<TTS-free>	on Egyptian Mummies
437	tourist		Is it free as well,
439	tourist		or do we need to pay an entrance fee for that?
455	SpaceBook	<TTS-free>	The special exhibition costs 7 pounds
460	tourist		I see.
464	tourist		Okay.
466	SpaceBook	<TTS-free>	But the museum is free to enter
468	tourist		Okay.
474	tourist		Opposite the museum.
476	SpaceBook	<TTS-button>	Is there anything else I can do for you?
478	tourist		Yes, opposite the museum I have spotted a dog statue.
483	tourist		May I know what it is,
485	tourist		like, about?
494	SpaceBook	<TTS-button>	I'm just finding that out for you now
509	SpaceBook	<TTS-free>	That is a statue of Greyfriar's Bobby
513	tourist		Can you tell me more about its history?
517	SpaceBook	<TTS-sequence>	The memorial to greyfriar's bobby lies on the sharp junction of Geroge Fourth Bridge and Candlemaker Row

Combined XML data output


```
- <Geo>
    <Latitude>55.94423562</Latitude>
    <Longitude>-3.18836916</Longitude>
    <Bearing>204.1</Bearing>
    <GTime>2011-08-23T17:06:37.971497+01:00</GTime>
  </Geo>
- <Geo>
    <Latitude>55.94421752</Latitude>
    <Longitude>-3.18837859</Longitude>
    <Bearing>188.1</Bearing>
    <GTime>2011-08-23T17:06:38.9647417+01:00</GTime>
  </Geo>
- <Message>
    <MessageSender>Wizard</MessageSender>
    <MessageString>TEXT_TO_SPEECH OK cotinue in the direction you are going</MessageString>
    <MTime>2011-08-23T17:06:39.1377516+01:00</MTime>
  </Message>
- <Geo>
    <Latitude>55.94420054</Latitude>
    <Longitude>-3.18837687</Longitude>
    <Bearing>168.1</Bearing>
    <GTime>2011-08-23T17:06:39.9377974+01:00</GTime>
  </Geo>
- <Geo>
    <Latitude>55.94418403</Latitude>
    <Longitude>-3.18837068</Longitude>
    <Bearing>168.1</Bearing>
    <GTime>2011-08-23T17:06:40.9272137+01:00</GTime>
  </Geo>
- <Geo>
    <Latitude>55.94416973</Latitude>
    <Longitude>-3.1883644</Longitude>
    <Bearing>164.8</Bearing>
    <GTime>2011-08-23T17:06:41.9428606+01:00</GTime>
  </Geo>
```



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