D6.1.2: Final Data Release, Wizard-of-Oz (WoZ) experiments

Robin Hill, Jana Götze, Bonnie Webber

Distribution: Public

SpaceBook

Spatial & Personal Adaptive Communication Environment: Behaviors & Objects & Operations & Knowledge

270019 Deliverable 6.1.2

28 February 2013

Project funded by the European Community under the Seventh Framework Programme for Research and Technological Development

The deliverable identification sheet is to be found on the reverse of this page.

Project ref. no.	270019
Project acronym	SpaceBook
Project full title	Spatial & Personal Adaptive Communication Environment: Behaviors & Objects & Operations & Knowledge
Instrument	STREP
Thematic Priority	Cognitive Systems, Interaction, and Robotics
Start date / duration	01 March 2011 / 36 Months

Security	Public
Contractual date of delivery	M24 = February 2013
Actual date of delivery	28 February 2013
Deliverable number	6.1.2
Deliverable title	D6.1.2: Final Data Release, Wizard-of-Oz (WoZ) experiments
Type	Report
Status & version	Final 1.0
Number of pages	18 (excluding front matter)
Contributing WP	6
WP/Task responsible	UEDIN
Other contributors	KTH, HWU
Author(s)	Robin Hill, Jana Götze, Bonnie Webber
EC Project Officer	Franco Mastroddi
Keywords	

The partners in SpaceBook are:

Umeå University	UMU
University of Edinburgh HCRC	UE
Heriot-Watt University	HWU
Kungliga Tekniska Högskola	KTH
Liquid Media AB	LM
University of Cambridge	UCAM
Universitat Pompeu Fabra	UPF

For copies of reports, updates on project activities and other SPACEBOOK-related information, contact:

The SPACEBOOK Project Co-ordinator:
Dr. Michael Minock
Department of Computer Science
Umeå University
Sweden 90187
mjm@cs.umu.se
Phone +46 70 597 2585 - Fax +46 90 786 6126

Copies of reports and other material can also be accessed via the project's administration homepage, `http://www.spacebook-project.eu`

1 Introduction

SpaceBook's Wizard-of-Oz (WoZ) experiments were aimed at determining pedestrian behaviour in interactive spatial navigation and exploration scenarios. First, two early-stage sets of experiments were carried out in Edinburgh:

Phase 1 experiments involved no special-purpose software or hardware: A mobile Wizard followed the pedestrian in the street; the Wizard therefore directly shared the same visibility and environmental conditions as the pedestrian, while interacting via a standard mobile phone line. Nine dyads were collected in this phase, referred to as *dyad1* through *dyad9*.

Phase 2 experiments placed the Wizard remotely in a room, interacting with the Pedestrian by voice via a telephone while monitoring their position on a computer screen using GPS tracking. Eight dyads were collected in this phase, labelled *dyad10* through *dyad17*.

The third phase of experiments (Phase 3) was supported by a bespoke GUI developed by Liquid Media that allowed the remote Wizard to track the pedestrians position while listening to them. Versions of this GUI (called the *WoZ Tool*), tailored to each task, underpin the final 15 dyads collected in Edinburgh, labelled *dyad18* through *dyad32* (Section 2), and all nine dyads collected in Stockholm (Section 3). From an interactional perspective, the key feature of these dyads is that the Wizard's interaction with the subject (i.e., the pedestrian) was limited to (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]*, or through the manual addition of free text on either side to form complete utterances, as with the sequence *Caffe Lucano*, to which the Wizard could manually append something like *"is situated near the museum"*; or (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).

Partially constraining the Wizard's contributions to the interaction in this way was not only intended to ease the Wizard's task and speed up the response time, it was intended to reduce subjects' tendency to "chatter" and attempt to engage their interlocutor (here, the Wizard) in common phatic interaction, a form of interaction that would interfere with SpaceBook's intended navigational and informational communicative goals. Comparing the Phase 3 dyads with the earlier ones shows a reduction in (but not complete elimination of) subject chatter.

A picture of the GUI used in Phase 3 of the Edinburgh WoZ experiments is shown in Figure 1, with a picture of that used in the complementary Stockholm WoZ experiments shown in Figure 3. Twin screens are used with live position displayed on one screen, and the TTS interface on the other. Note that related utterances are grouped together in a single row (e.g., *Keep going straight ahead*, *Take the next left*, *Turn at the next right* in Figure 1, and *Turn right*, *Turn right into englebreksgatan* in Figure 3), quickly clarifying the preset options. And as much as possible, the order of the rows reflects when in the dialogue their options are likely to be relevant.

2 Edinburgh WoZ Experiments

The purpose of the Edinburgh WoZ experiments was to characterize how pedestrians will use SpaceBook to help them navigate and explore urban space specifically, to characterize: (1) the type of utterances they

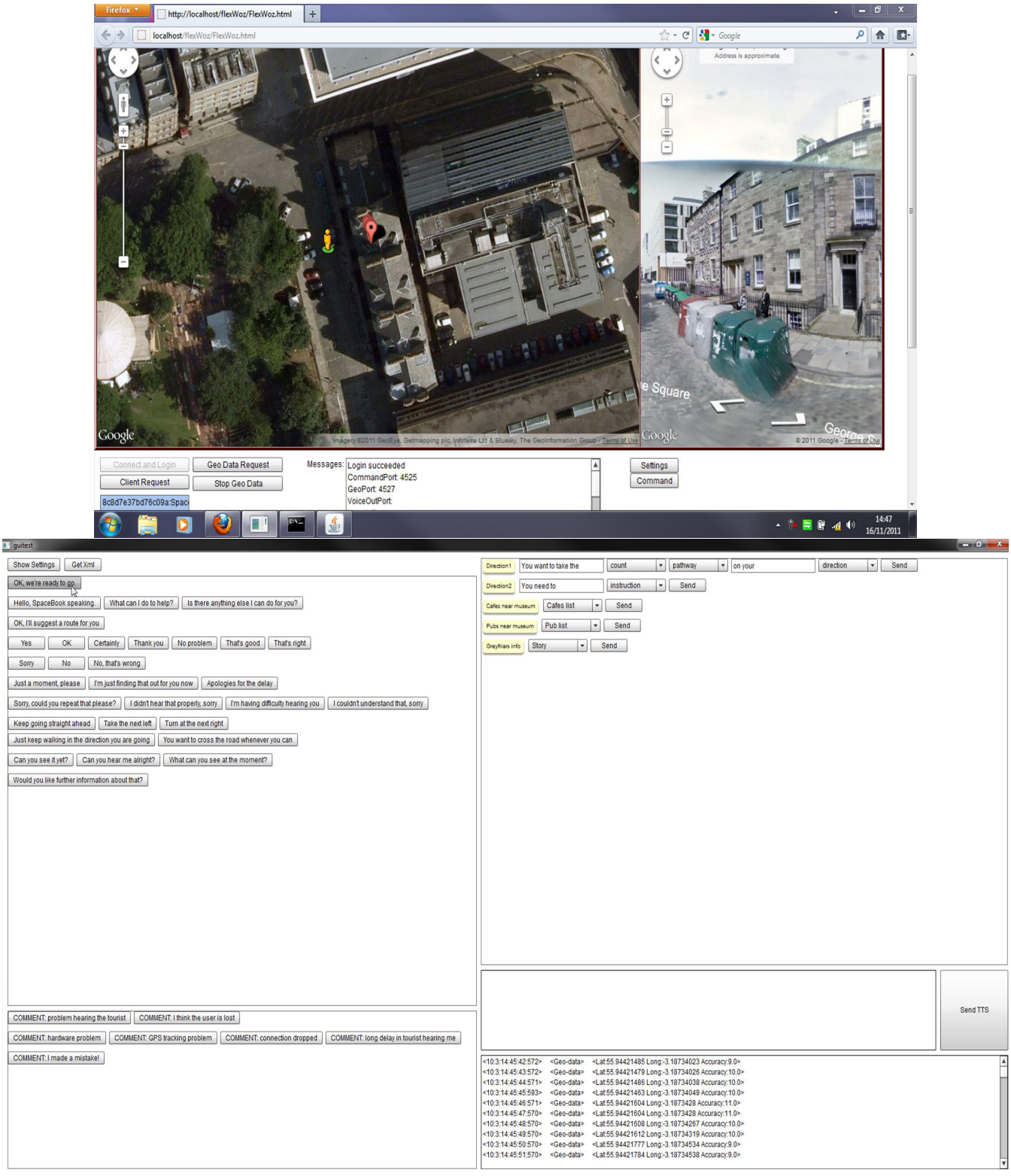


Figure 1: WoZ interface, as seen during the collection of Dyad 19, during phase 3 data collection. Incorporated images © 2011 Google and © 2011 GeoEye.

Dyad ID	Interaction Length (minutes)	Number of Turns S	Number of Turns W	Wizard TTS button/seq/free	Note
18	17:18		60	40/12/8	Ended early
19	45:28		105	4/0/101	
20	49:48		136	56/2/78	
21	13:38		32	17/3/12	Ended early
22	19:58		75	16/6/53	
23	25:38	65	41	15/3/23	
24	9:13	12	23	10/0/13	Ended early
25	28:12	85	83	25/5/53	
26	23:46	148	68	16/0/52	
27	33:47	87	98	32/7/59	
28	20:07	44	80	10/0/70	Only have WoZ utterances for final 9 minutes
29	28:45	34	126	38/12/76	
30	65:50	221	239	58/14/167	Extended discussion of restaurants & library
31	21:24	112	111	38/11/62	
32	54:34	405	143	12/0/131	

Table 1: Summary of the Phase 3 audio transcriptions. S–Subject, W–Wizard.

will use with SpaceBook; (2) the specific kinds of things they will refer to and how they will refer to them; (3) their frequency of interaction with SpaceBook; and (4) the ways they accommodate to SpaceBooks abilities and communicative behavior, which differs from both normal human communication and simple formulaic computer-human phone interaction.

The scenario was the same in all experiments carried out in Year 1 (as laid out in Appendix B, Deliverable D6.1.1). It was meant to confront the pedestrian and the Wizard with a range of environments (a complex street system – as in the path to the National Museum of Scotland – as well as an open area – as in the Meadows, cf. Figure 2), with both navigational and informational tasks, and both specific and vague goals.

In total, 32 dyads were captured and transcribed. Because TTS was adopted for Wizard utterances in Phase 3, transcription of Phase 3 dyads only required work on the subject's utterances. In the transcripts being released to the public, Wizard text strings have been interleaved with the transcribed subject utterances to produce full dyads. In three of the dyads (dyad 23, 25 and 30), this was done manually, and both timings and Wizard utterance types (button-based *TTS-button*, sequence-based *TTS-sequence*, or free text *TTS-free*) are provided. The other twelve were done automatically and both timing and utterance type are missing. The transcription itself was done by HWU in Edinburgh using its adaptation of the ELAN tool.

Here we provide a brief analysis of the Phase 3 dyads from the Edinburgh transcripts that elaborates on the figures presented in Table 1. This is primarily an analysis of the Wizard's utterances, as detailed analysis has not yet been done on those of the subjects.

With respect to the Wizard's use of fixed strings (*buttons*), although the dyads varied significantly in the

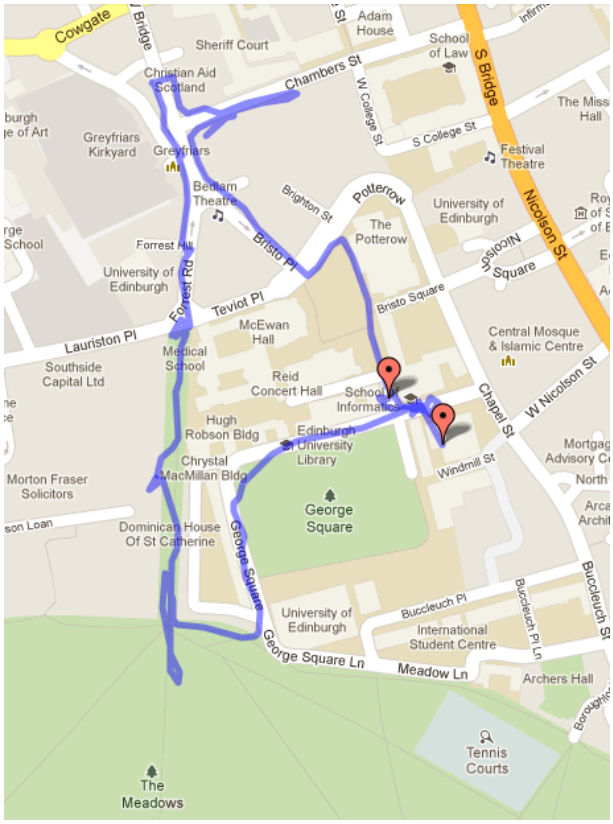


Figure 2: Route followed in scenario used in Edinburgh WoZ experiments. Map data © 2011 Google.

number of times they were used in the dialogue, the distribution of Wizard message types is of interest. Recall that each row of buttons contain related messages. Of the 398 tokens of fixed messages used in dyads 18-32, 28.6% (114) were *confirmations* (Row 4, comprising *yes*, *OK*, *certainly*, *Thank you*, *No problem*, *That's good*, and *That's right*). After that came messages indicating a delay in responding (12.1%, 48 tokens), comprising *Just a moment please*, *I'm just finding that out for you now* and *Apologies for the delay*, and messages responding to problems in understanding (13.3%, 53 tokens), including *I couldn't understand that, sorry* and *I didn't hear that properly, sorry*. Fixed strings related to navigation comprised only 18.3% (73 tokens) of the Wizard's utterances, including *Keep going straight ahead* and *Just keep walking in the direction you are going*.

With respect to the Wizard's use of modifiable text strings (*sequences*), we can distinguish between *Named Entities* (pubs and cafes) that are incorporated into other text strings, *templates*, which have a variable that needs to be filled in through selection or free choice, and *sentence elements* from multi-sentence descriptions of a touristic attraction. Of the 75 sequences that the Wizard made use of, 15 were named entities, 40 were story entries, and 20 were templates, of which 16 were *You need to X* and four were *You want to X*, where X is a free text string.

Finally, with respect to the Wizard's use of free text, strings are counted as separate "utterances" if they were produced from distinct calls to TTS. Thus a single sentence (in the grammatical sense) may be counted as several distinct utterances as in

> The gravestone is near the entrance
> on the right
>
> The exhibition is called
> Watching on the Heights

which are both analysed as composed of two utterances. While we have not yet done a complete analysis of the Wizard's use of free text, it is clear that in a range of cases, he could have used fixed strings (*buttons*) or modifiable text strings (*sequences*). The point here is that a greater percentage of the Wizard's utterances could have been covered by fixed or modifiable text strings than is shown by the number of free text utterances.

3 Stockholm WoZ Experiments

This section presents a Wizard-of-Oz (WoZ) study conducted in Stockholm. The purpose of the study was twofold: Firstly, to investigate which kinds of route-giving instructions seem to be misunderstood and/or lead to clarification questions from the subject in certain geographic situations; secondly, to investigate which kinds of clarification questions and other feedback that subjects tend to provide. The first item is useful for dialog modeling for such route-giving dialog systems, both from the system and the (simulated) user point-of-view; the second item is useful for the purpose of language modeling and parsing.

The WoZ experiment involved a person (henceforth called the subject) receiving instructions to walk around in the city of Stockholm in a treasure-hunt scenario. The subjects were told that they will receive instructions from an automated system and were asked to converse with it as they wished. In reality, the experimenter (the Wizard) acted as the system.

3.1 Apparatus

The subject was equipped with an Android mobile phone (Motorola Razr). The phone ran the SpaceBook Application (Deliverable 5.1), which allows to record the subject's GPS data and speech signal. It also allows the Wizard to send messages to the subject via text-to-speech (TTS).

The Wizard sat in a laboratory, using the WoZ interface described in D5.1. This allowed him to see the subject's position on a map and type messages that were sent to the subject via TTS. Figure 3 shows the WoZ interface before an experiment.

3.2 Procedure

Participants were randomly selected from a number of PhD students who volunteered to participate. Participants signed a consent form (Appendix B) and provided demographic information (age, sex, native language, familiarisation with the geographical area) and a short questionnaire after the experiment (Appendix C). The participants were orally informed about their task in the experiment, and the functionality of the system (cf. Appendix D).

The subject was then sent to go to the starting point of the experiment, which was a few hundred meters away from the laboratory. During their walk there, the Wizard sent some example messages like "You are doing fine", or "There is an X to your left" so that the subject could accommodate to how the synthesized speech sounds and to what the system knows about their position.

If the connection to the phone dropped for any reason, the subject would re-start the application and continue. The subject was told to find some objects in the city and write down something that the system would ask them along the way. The subject was told that the system would give them instructions and their task was to follow them as well as they could and that they could ask for clarification in any way that they wanted.

3.3 Route instructions and WoZ interface setup

The Wizard's task was to guide the subject along a fixed route to three places in the city where they should find their way to an object that is part of the environment, but that is not immediately conspicuous, e.g. an electricity box in the street, as shown in Figure 4. The WoZ interface was programmed to contain buttons for a number of previously determined instructions that were to be given at certain way points (cf. Figure 5) and some general phrases. As soon as the subject reached a certain point on the map, the Wizard used the corresponding button.

The buttons on the interface as seen in Figure 3 are ordered in the sequence that they were to be given along the route. In order to simplify the task for the Wizard, buttons are repeated if they had to be pressed again at a later point. In that way, the Wizard was able to go through the list downwards without having to return to an earlier item (note that this excludes the general phrases in the top of the interface which could be used at any point during the dialog).

Some longer instructions are split up into shorter utterances and the buttons had to be pressed one after the other, e.g. "Walk into Östermalmsgatan", "The school of architecture is behind you". Similarly, where a short instruction like "Turn right" could potentially cause confusion, a longer version that included more information was also provided, e.g. "Turn right into Engelbrektsgatan". Splitting up phrases and providing shorter versions of some instructions also turned out to be useful for giving repetitions or clarifications,

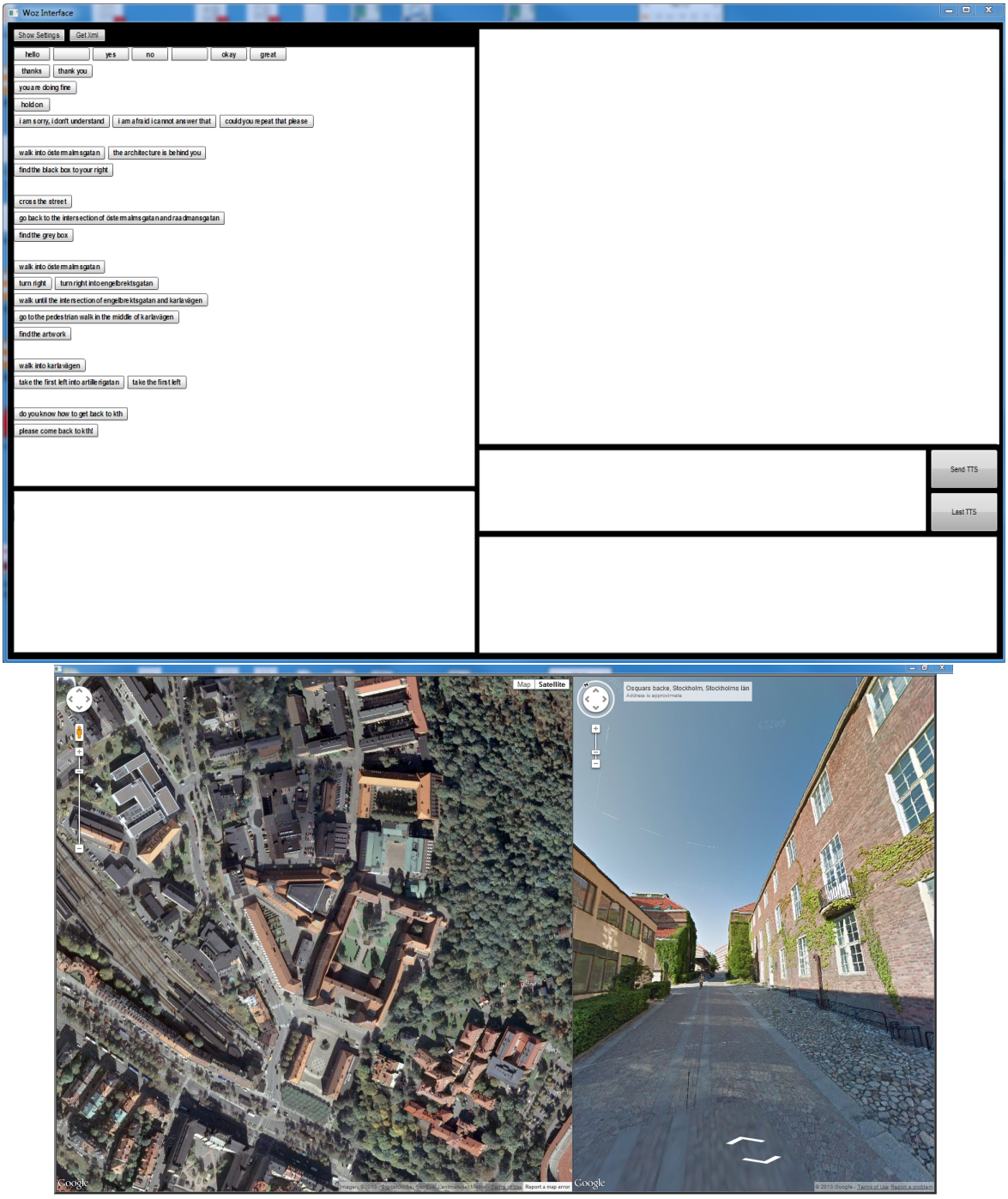


Figure 3: The WoZ tool containing buttons that the Wizard used to send text-to-speech messages to the subject and the map view of where the subject is according to his GPS coordinates. Incorporated images © 2013 Google and © 2013 GeoEye.

Figure 4: A black electricity box that the subject was asked to find. Incorporated images © 2012 GeoEye.

where the subject sometimes asked for a particular piece of information to be repeated (instead of the whole instruction).

There were also buttons for general phrases like "Yes", "No", "You are doing fine", "Please turn around" and "Sorry, I didn't understand, could you please repeat". If necessary, the Wizard could also type free text. This feature was however to be avoided as much as possible. The Wizard's task was to constrain himself as much as possible to the current functionality of a route-giving system, which mostly consisted in giving, repeating, and rephrasing instructions and answering simple queries, such as "Should I turn left here?".

In order to elicit clarification questions from the subjects that could later be helpful for language modeling and dialogue management purposes, some instructions were intentionally constructed ambiguously or contained wrong information. For example, instructions 6 and 11 in Figure 5 omit the direction when asking the subject to turn into a street X, and instruction 12 states a street name that conflicts with the rest of the instruction.

3.4 Data generated

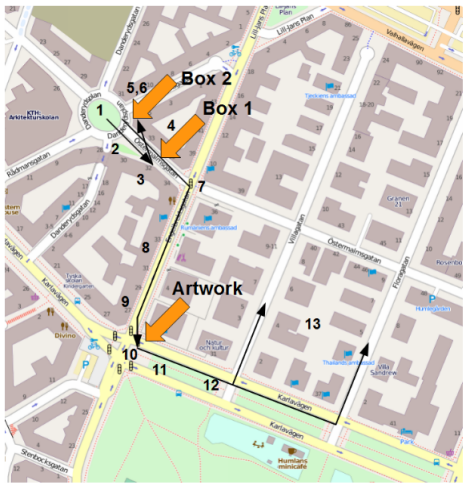
The subject's speech was recorded using Audacity[1], his GPS position was logged every second through the SpaceBook phone application. The Wizard's TTS messages were logged. All log files are time-stamped and were synchronized accordingly.

3.4.1 Data processing

The subjects' speech was transcribed and annotated using the Higgins Annotation Tool[2]. Each utterance is annotated with the speaker, and, for those of the Wizard, an indication of whether the speech came from

[1] http://audacity.sourceforge.net/

[2] http://www.speech.kth.se/hat/



Minimal set of instructions:

1) Walk into Östermalmsgatan. The school of architecture is behind you.
2) Find the black box to your right.
3) Cross the street.
4) Go back to the intersection of Östermalmsgatan and Rådmansgatan.
5) Find the grey box.
6) Walk into Östermalmsgatan.
7) Turn right.
8) Walk until the intersection of Engelbrektsgatan and Karlavägen.
9) Go to the pedestrian walk in the middle of Karlavägen.
10) Find the artwork.
11) Walk into Karlavägen.
12) Take the first left into Artillerigatan.
13) Please come back to KTH.

Figure 5: The route with objects and the instructions used by the Wizard to guide the subject. (Map: ©OpenStreetMap contributors, `http://www.openstreetmap.org`)

a TTS button (i.e. a pre-determined phrase) or was freely typed.
Table 2 shows some statistics on the length of the dialogs and the number of utterances and turns. In the case of the Wizard speech, an utterance boundary was inserted for every message that was sent through the TTS. The subjects' speech was segmented into utterances at either phrase boundaries or when there was a pause of 2 seconds or more.

3.4.2 Data

Nine subjects were run, in total. The area of Stockholm in which the experiment was carried out was close to the subjects' workplace. In the questionnaire, eight subjects reported familiarity with the area around the first two objects they had to find (these two electrical were very close to each other, cf. Figure 5). For the third object, an artwork, only two reported to be familiar with the area.

Seven subjects found all three objects, one did not find the first box and one could only find the artwork. There were no strict rules for the Wizard as when to give up on one object and proceed to the next, this decision was to be made depending on the situation with the conditions that the subject should not get too frustrated and the Wizard should not deviate too much from the given phrases, i.e. try to avoid using many free text TTS utterances.

Table 2 shows that the dialogs lasted between 13.8 and 19.3 minutes and that the turns that each, subject and Wizard, took ranged from 15 to 51 where these two measures do not seem to be related, but some subjects were more talkative than others. Subject 3 was an extreme case that took the turn 51 times and was actively taking the initiative in the dialog and asking many clarification questions. On the other end of the scale was subject 6, who mostly only gave short acknowledgments or repetition requests. In contrast, the latter dialog was slightly longer than the first.

Some of the instructions that the Wizard provided were intentionally left ambiguous. In two cases, the direction of an instruction was left out in the assumption that the subject might be able to infer it from the preceding actions:

Subject ID	Interaction Length in minutes	Number of Utterances S	W	Wizard TTS Number of button/free text*	Number of Turns S	W	Objects found
1	19:18	47	52	26/26	40	40	1
2	14:24	52	45	22/23	38	37	3
3	15:36	123	57	19/38	51	51	3
4	13:48	48	44	20/24	39	38	2
5	14:00	17	40	22/18	16	17	3
6	16:30	17	39	33/6	15	15	3
7	19:36	47	48	37/11	39	38	3
8	15:24	23	36	27/9	21	21	3
9	13:54	24	39	32/7	23	24	3
Average	15:48	44.2	44.4	26.4/18	31.3	31.2	

Table 2: Summary of the audio transcriptions. S–Subject, W–Wizard. (*)Note that the number of fixed TTS buttons changed over the course of the experiment.

- In "Walk into Östermalmsgatan" (instruction 6 in Figure 5) the subject had just come from the direction that he is now intended to go to and therefore he could either continue in the direction that he is walking or go back to a street in which he had already been.
  In the eight dialogs where this situation was applicable, 5 subjects asked for clarification, e.g. "What direction?", and 3 did not ask for clarification, 2 of them walked in the intended direction and 1 subject walked in the opposite direction.

- In "Walk into Karlavägen" (instruction 11), the subject has already started walking in the intended direction (by finding the third object), but it has not been explicitly mentioned. In this situation, the assumption is that the intended choice will be more salient than in the first example, where both possible choices would make sense.
  Here, more subjects seemed to be confident that they were making the correct choice, five of nine walked in the intended direction without asking for clarification. Three subjects asked for clarification and accepted "Walk further into Karlavägen" as an answer and one walked in the opposite direction without asking.

- In "Take the first left into Artillerigatan" (instruction 12), the name of the street is in conflict with the instruction "Take the first left". This instruction was intended to elicit a clarification from the subjects, because they were explicitly told that the system knew where they were.
  Only four of the 9 subjects openly asked for clarification when they noticed the conflict. Two chose to ignore the name of the street and took the first possible left turn, and three chose to ignore the information "first turn" and continued walking to look for the correct street. These latter three were then instructed to "Take the first left".

All other Wizard instructions were intended to give straightforward information and hardly caused confusion. Although all subjects reported that they had trouble understanding the TTS voice when it was pronouncing Swedish street names, the number of requests for repetitions was low, two subjects asked 4

times each, all others asked at most once.

Looking at the types of utterances that the subjects use, the largest part consists of acknowledgments like "okay" (ca. 19.5%) and informs like "I am walking along street X now" (ca. 24.5%). Request, e.g. for the next instruction or for a clarification, made up ca. 11.7%.

3.5 Discussion

This data collection aimed at investigating two specific things: (1) the kinds of route-giving instructions that the subject seemed to misunderstand or that led them to ask clarification questions in certain geographic situations, and (2) the kinds of clarification questions and other feedback that subjects provided.

The first point is useful for developing dialog management behavior. We investigated two different situations: (1) where the system does not give enough information, as in the first two examples in Section 3.4.2, and (2) where the system gives incorrect information, as in the third example. The first case can arise when the system lacks sufficient information on the subject's bearing to include directional information (e.g. left or right) or it does not recognize that its instruction is ambiguous, with another possible street that the subject could choose. The second case can arise when there are errors in the geographical data. It is therefore useful to know how a subject might react in such a situation.

While the first type of case leads most of the subjects to request clarification when they notice the ambiguity, which can simply be answered, the data highlight the need for the dialog manager to address the second type of case. This is because half of the subjects do not signal a problem using speech and just move along the path that they think is correct.[3] Where this is not the system's intended route, it might be useful for the dialog manager to be able to adapt its strategy according to its explanation for the subject's behavior, ranging from not hearing the system, to trying to make sense of it, to ignoring the instruction.

The second point, the kinds of clarifications and feedback the subjects use, is useful for the purpose of language modelling and parsing. To this end, it is interesting to look at how the subjects phrased their requests for clarification, so that the semantic parser described in D4.1.1 can be further extended.

4 Conclusion

The data collected in these experiments in Edinburgh and Stockholm have been valuable in shaping the design of the SpaceBook system. With more resources, additional experiments would have been worth carrying out.

[3]Note that subjects also report difficulties with the TTS pronunciation of street names, which also may have played a role here, but cannot be seen in the data.

Appendix A: Transcription Guidelines

Spacebook transcription guidelines[1]

General Guidelines

Basic spelling is British English, i.e., "colour" vs. "color", "realise" vs. "realize".

1. Each segment should be padded by a small buffer of silence (1/4 to 1/2 second) on both sides (before and after the speech) if possible. If there are two or more short utterances with only a small gap between them, consider combining them into one longer one.

2. It is not generally necessary to record noises such as coughs and laughs. However, if you feel that a noise really needs to be recorded, it is best to add a note to the 'meta' tier.

3. Transcribe verbatim, without correcting grammatical errors, e.g., "I seen him", "me and him have done this".

4. Standard spoken language should be transcribed as it is spoken, e.g., "gonna" not "going to", "wanna" not "want to", "kinda" not "kind of", etc. See the table of regularised spellings below for more examples.

5. Avoid word abbreviations, i.e., "doctor" not "Dr", and "mountain" not "Mt".

6. Use normal capitalization on proper nouns and at the beginning of sentences.

7. Remember to watch for common spelling confusions like "its" and "it's", "they're" and "there" and "their", "by" and "bye", "of" and "off", "to" and "too", etc.

8. Spell out number sequences, e.g., "forty four" not "44". Please transcribe *without* a hyphen.

9. Acronyms should be spelt as they are pronounced, e.g., "NASA" or "U_S_A_".

10. When a letter sequence is used as part of a word, add the inflection after the underscore: "They I_D_ed him".

11. If a speaker does not finish a word, and you think you know what the word was, you can spell out as much of the word as was pronounced inserting a single dash as the last letter of the word, e.g., "I think basic- ".

12. If the speaker is cut off or doesn't finish their sentence, a dash should be inserted at the end of the statement, e.g., "I was going to do that, but then -" (remember, if there is no space between the last word and the dash, this indicates that the *word* was not finished). If the utterance ends in a word fragment, it is not necessary to add a separate dash to indicate that the sentence was also left unfinished.

13. Punctuation should be limited in the corpus. You should only use commas, full stops (periods), and question marks to punctuate a 'sentence'. You can include dashes in compound words that are traditionally written with them, e.g., 'passer-by', but make sure

[1] Based on guidelines developed and used in the AMI (http://www.cstr.ed.ac.uk/research/projects/ami/) and JAST (http://www.ltg.ed.ac.uk/projects/JAST) projects.

there is no space between the dash and the words on either side of it. Please don't use colons, semicolons, or quotation marks. If you come across a phrase that you feel needs to be offset, e.g., a parenthetic remark, please use commas.

14. Particular care should be taken at two locations:

    a. at regions of disfluency, where the speaker interrupts himself to correct or restart or repeat, use a dash (if there is not already a dash from a word fragment), e.g., "I just meant - I mean"; and

    b. at the end of a speech segment, punctuation should be used to make it clear whether the next turn is a continuation of the current one. If the speaker continues with the same utterance, punctuate as you would if there was no break (including potentially having no punctuation at all at the end of the initial turn); if the speaker breaks off and does not continue the sentence in his next turn, then indicate this with a dash.

15. Mispronunciations: if a speaker mispronounces a word and you know what word was intended, transcribe the word as it should be spelled and mark it with an asterisk after the last letter, e.g., "spaghetti*". If you do not know what word was intended, transcribe what you hear and mark it with parentheses, e.g., "(fligop)". If you have no idea at all, use "(??)".

List of Regularised Spellings

These are the spellings to use for common non-words or tricky to spell words.

CONTRACTIONS	gonna, wanna, gotta, kinda, sorta, shoulda, woulda, coulda, dunno, lotta (= 'lot of')
AGREEMENT	uh-huh, mm-hmm, yeah, yep, aye (Scottish)
DISAGREEMENT	uh-uh, mm-mm, nope, nah
BACKCHANNELS	ah, huh, hmm, mm
HESITATIONS	uh (any variant of a pure vowel), um (any variant of vowel plus nasal), mm (any variant of pure nasal)
TAG QUESTIONS	eh, huh
INITIAL ELLIPSIS	'kay (= okay), 'scuse (= excuse), 'til (= until), 'cause (= because), 'em (= them)
ACRONYMS AND PROPER NAMES	L_C_D_, T_V_, R_B_S_, iPod, A_SAP (pronounced 'ay-sap')
INTERJECTIONS/ EXCLAMATIONS	ah, ah-ha, argh, doh, gee, geez, oh, ooh, phew, whoa, whoo, whoohoo, yay

Appendix B: Consent Form used in Stockholm Experiments

Informed Consent Form for Experimental Participants

Please read the following information carefully. You can also request a copy for future reference.

Experiment: City-Navigation using the SpaceBook Spoken Dialog System
Experimenters: Johan Boye and Jana Götze
Affiliation: KTH Royal Institute of Technology

DESCRIPTION: You are invited to participate in a research study that investigates route-giving and route-following behavior in human-machine interaction. You will interact with a spoken dialog system that gives you instructions on what to do. Your speech and GPS position will be recorded so that we can study how people interact with the system in different situations. You will also be asked to fill out a questionnaire afterwards.

RISKS AND BENEFITS: You will carry a smartphone that records speech and position data about your actions during the experiment. There are no known risks involved in this procedure.

TIME INVOLVEMENT: Your participation will take approximately 90 minutes.

SUBJECT'S RIGHTS: If you have read this form and have decided to participate in this experiment, please understand your participation is voluntary and you have the right to withdraw your consent or discontinue participation at any time without penalty. You have the right to refuse to answer particular questions. Your individual privacy will be maintained in all published and written data resulting from the study. Note that your voice is recorded and may thus be recognizable.

If you agree with the above-stated conditions and are willing to participate in the experiment, please sign below. By signing the form, you confirm that you meet the following conditions:

- You have read the above consent form, understood it and you agree to it.
- You want to participate in the above-mentioned experiment.

Name:

Date:

Signature:

Appendix C: Subject Questionnaire used in Stockholm Experiments

- Age:
- Sex:
- Native Language:
- Were you familiar with the areas
  - around the electrical boxes? yes/no
  - around the artwork? yes/no
- Did you think it was easy to follow the instructions? yes/sometimes/no
- Were the instructions always clear? If not, elaborate.
- Were you confident in your actions? yes/sometimes/no
- Did you feel that the system understood you? always/mostly/rarely/never
- Do you think that the system could have talked more? If yes, elaborate.
- Did you feel that the system knew where you were? Elaborate.

Appendix D: Description of Task used in Stockholm Experiments

Subjects were given the following instruction when starting the experiment:

The system will guide you in the streets and sometimes ask you to find things along the way. The experiment will start when you are in Danderydsplan. Try to follow the instructions as good as you can. There are no special commands for the system, you can speak as you want to it, e.g. if you don't understand or need more information.