Evaluating a City Exploration Dialogue System Combining Question-Answering and Pedestrian Navigation

Srinivasan Janarthanam[1], **Oliver Lemon**[1], **Phil Bartie**[2], **Tiphaine Dalmas**[2],
Anna Dickinson[2], **Xingkun Liu**[1], **William Mackaness**[2], and **Bonnie Webber**[2]

[1] The Interaction Lab, Heriot-Watt University

[2] Edinburgh University

`sc445@hw.ac.uk`

Abstract

We present a city navigation and tourist information mobile dialogue app with integrated question-answering (QA) and geographic information system (GIS) modules that helps pedestrian users to navigate in and learn about urban environments. In contrast to existing mobile apps which treat these problems independently, our Android app addresses the problem of navigation and touristic question-answering in an integrated fashion using a shared dialogue context. We evaluated our system in comparison with Samsung S-Voice (which interfaces to Google navigation and Google search) with 17 users and found that users judged our system to be significantly more interesting to interact with and learn from. They also rated our system above Google search (with the Samsung S-Voice interface) for tourist information tasks.

1 Introduction

We present a mobile dialogue system (an Android app) called **Spacebook** that addresses the problem of pedestrian navigation and tourist information in urban environments. There has been little prior work that addresses these two problems - navigation and tourist information provision - in an integrated way. By *navigation*, we refer to the problem of finding appropriate destinations to go to and the task of wayfinding to reach them and by *tourist information provision* we refer to the problem of meeting the informational needs of a user about entities such as museums, statues and famous personalities. A dialogue system such as this could serve as a personal tour guide to pedestrian tourists as they walk around unknown cities. With the proliferation of smartphones, there has been a

number of mobile apps developed to address these problems. However these apps have the following problems: first, they demand the user's visual attention because they predominantly present information on a mobile screen. This can be dangerous in urban environments, as well as being distracting. Second, these apps address the problems of navigation and tourist information independently and therefore do not have a shared interaction context. This means that users cannot switch between information and navigation tasks in a natural and fluid manner.

User1: Take me to the National Museum.
System2: The National Museum is about 300m away..
System3: At the KFC, turn left on to South Bridge
System4 : Near you is the statue of David Hume.
User2: Who is David Hume.
System5: David Hume was a Scottish philosopher....
User3: Tell me more about David Hume.
System6: He was one of the most important figures in..
System7: You should be able to see **the museum** ...
User4: Tell me more about **the museum**.
System8: **The National Museum of Scotland** is a....

Table 1: An example interaction with the evaluated system

In contrast to many existing mobile apps, Spacebook has a speech-only interface and addresses both problems in an integrated way. We conjecture that with a speech-only interface, users can immerse themselves in exploring the city, and that because of the shared context they can switch between navigation and tourist information tasks more easily. Using the navigational context, Spacebook pushes point-of-interest information which can then initiate tourist information tasks using the QA module. Table 1 presents an example interaction with our system showing the integrated use of navigation and question-answering capabil-

ities. Utterances *System4-8* show the system's capability to push information about nearby points-of-interest (PoI) during a navigation task and answer followup questions using the QA system (in utterances *User2* and *User3*). The final 3 utterances show a natural switch between navigation to an entity and QA about that entity.

We investigate whether our system using a combination of geographical information system (GIS) and natural language processing (NLP) technologies would be a better companion to pedestrian city explorers than the current state-of-the-art mobile apps. We hypothesize that, (1) users will find our speech-only interface to navigation efficient as it allows them to navigate without having to repeatedly look at a map and (2), that users will find a dialogue interface which integrates touristic question-answering and navigation within a shared context to be useful for finding information about entities in the urban environment. We first present some related work in section 2. We describe the architecture of the system in section 3. We then present our experimental design, results and analysis in sections 5, 6 and 7.

2 Related work

Mobile apps such as Siri, Google Maps Navigation, Sygic, etc. address the problem of navigation while apps like Triposo, Guidepal, Wikihood, etc. address the problem of tourist information by presenting the user with descriptive information about various points of interest (PoI) in the city. While some exploratory apps present snippets of information about a precompiled list of PoIs, other apps dynamically generate a list of PoIs arranged based on their proximity to the users. Users can also obtain specific information about PoIs using Search apps. Also, since these navigation and exploratory/search apps do not address both problems in an integrated way, users need to switch between them and therefore lose interaction context.

While most apps address these two problems independently, some like Google Now, Google Field Trip, etc, mix navigation with exploration. But such apps present information primarily visually on the screen for the user to read. Some of these are available for download at the Google Play Android app store[1]. Several dialogue and natural language systems have addressed the issue of pedestrian navigation (Malaka and Zipf, 2000; Raubal and Winter, 2002; Dale et al., 2003; Bartie and Mackaness, 2006; Shroder et al., 2011; Dethlefs and Cuayáhuitl, 2011). There has also been recent interest in shared tasks for generating navigation instructions in indoor and urban environments (Byron et al., 2007; Janarthanam and Lemon, 2011). Some dialogue systems deal with presenting information concerning points of interest (Ko et al., 2005; Kashioka et al., 2011) and interactive question answering (Webb and Webber, 2009).

In contrast, Spacebook has the objective of keeping the user's cognitive load low and preventing users from being distracted (perhaps dangerously so) from walking in the city (Kray et al., 2003). Also, it allows users to interleave the two sub-tasks seamlessly and can keep entities discussed in both tasks in shared context (as shown in Table 1).

3 Architecture

The architecture of the Spacebook system is shown in figure 1. Our architecture brings together Spoken Dialogue Systems (SDS), Geographic Information Systems (GIS) and Question-Answering (QA) technologies (Janarthanam et al., 2012). Its essentially a spoken dialogue system (SDS) consisting of an automatic speech recogniser (ASR), a semantic parser, an Interaction Manager, an utterance generator and a text-to-speech synthesizer (TTS). The GIS modules in this architecture are the City Model, the Visibility Engine, and the Pedestrian tracker. Users communicate with the system using a smartphone-based client app (an Android app) that sends users' position, pace rate, and spoken utterances to the system, and delivers synthesised system utterances to the user.
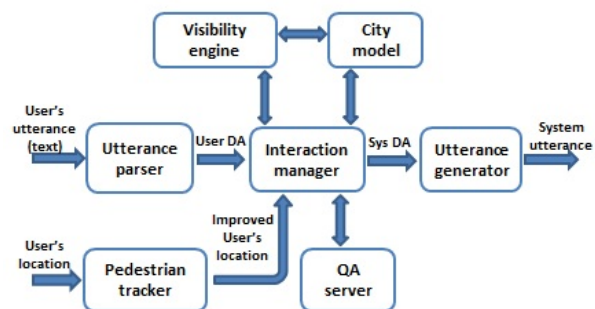

Figure 1: System Architecture

3.1 Dialogue interface

The dialogue interface consists of a speech recognition module, an utterance parser, an interaction manager, an utterance generator and a speech synthesizer. The Nuance 9 speech recogniser with a domain specific language model was used for speech recognition. The recognised speech is currently parsed using a rule-based parser into dialogue acts and semantic content.

The Interaction Manager (IM) is the central component of this architecture, which provides the user with navigational instructions, pushes PoI information and manages QA questions. It receives the user's input in the form of a dialogue act (DA), the user's location (latitude and longitude) and pace rate. Based on these inputs and the dialogue context, it responds with system output dialogue act, based on a dialogue policy. The IM initiates the conversation with a calibration phase where the user's initial location and orientation are obtained. The user can then initiate tasks that interest him/her. These tasks include searching for an entity (e.g. a museum or a restaurant), requesting navigation instructions to a destination, asking questions about the entities in the City Model, and so on. When the user is mobile, the IM identifies points of interest[2] on the route proximal to the user. We call this "PoI push". The user is encouraged to ask for more information if he/she is interested. The system also answers adhoc questions from the user (e.g. "Who is David Hume?", "What is the Old College?", etc) (see section 3.4).

Navigation instructions are given in-situ by observing user's position continuously, in relation to the next node (street junction) on the current planned route, and they are given priority if in conflict with a PoI push at the same time. Navigation instructions use landmarks near route nodes whenever possible (e.g. "When you reach Clydesdale Bank , keep walking forward"). The IM also informs when users pass by recognisable landmarks, just to reassure them that they are on track (e.g. "You will pass by Tesco on the right"). In addition to navigation instructions, the IM also answers users' questions concerning the route, his/her location, and location of and distance to the various entities. Finally, the IM uses the city model's Visibility Engine (VE) to determine whether the destination is visible to the user (see section 3.3).

[2]Using high scoring ones when there are many, based on tourist popularity ratings in the City Model.

The shared spatial and dialogue context employs a feature-based representation which is updated every 1 second (for location), and after every dialogue turn. Spatial context such as the user's coordinates, street names, PoIs and landmarks proximal to the user, etc are used by PoI pushing and navigation. The dialogue context maintains the history of landmarks and PoIs pushed, latest entities mentioned, etc to resolve anaphoric references in navigation and QA requests, and to deliver coherent dialogue responses. The IM resolves anaphoric references by keeping a record of entities mentioned in the dialogue context. It also engages in clarification sub-dialogues when the speech recognition confidence scores are low. The IM stores the name and type information for each entity (such as landmark, building, etc) mentioned in navigation instructions and PoI pushes. Subsequent references to these entities using expressions such as "the museum", "the cafe" etc are resolved by searching for the latest entity of the given type. Pronouns are resolved to the last mentioned entity.

The IM also switches between navigation, PoI push, and QA tasks in an intelligent manner by using the shared context to prioritise its utterances from these different tasks. The utterance generator is a Natural Language Generation module that translates the system DA into surface text which is converted into speech using the Cereproc Text-to-Speech Synthesizer using a Scottish female voice. The only changes made were minor adjustments to the pronunciation of certain place names.

3.2 Pedestrian tracker

Urban environments can be challenging with limited sky views, and hence limited line of sight to satellites, in deep urban corridors. There is therefore significant uncertainty about the user's true location reported by GNSS sensors on smartphones (Zandbergen and Barbeau, 2011). This module improves on the reported user position by combining smartphone sensor data (e.g. accelerometer) with map matching techniques, to determine the most likely location of the pedestrian (Bartie and Mackaness, 2012).

3.3 City Model

The City Model is a spatial database containing information about thousands of entities in the city of Edinburgh (Bartie and Mackaness, 2013). This data has been collected from a variety of exist-

ing resources such as Ordnance Survey, Open-StreetMap, Google Places, and the Gazetteer for Scotland. It includes the location, use class, name, street address, and where relevant other properties such as build date and tourist ratings. The model also includes a pedestrian network (streets, pavements, tracks, steps, open spaces) which is used by an embedded route planner to calculate minimal cost routes, such as the shortest path. The city model also consists of a Visibility Engine that identifies the entities that are in the user's *vista space* (Montello, 1993). To do this it accesses a *digital surface model*, sourced from Li-DAR, which is a 2.5D representation of the city including buildings, vegetation, and land surface elevation. The Visibility Engine uses this dataset to offer a number of services, such as determining the line of sight from the observer to nominated points (e.g. which junctions are visible), and determining which entities within the city model are visible. Using these services, the IM determines if the destination is visible or not.

3.4 Question-Answering server

The QA server currently answers a range of *definition* and *biographical* questions such as, "Tell me more about the Scottish Parliament", "Who was David Hume?", "What is haggis?", and requests to resume (eg. "Tell me more"). QA is also capable of recognizing out of scope requests, that is, either navigation-related questions that can be answered by computations from the City Model and dealt with elsewhere in the system ("How far away is the Scottish Parliament?", "How do I get there?"), or exploration queries that cannot be handled yet ("When is the cannon gun fired from the castle?"). Question classification is entirely machine learning-based using the SMO algorithm (Keerthi et al., 1999) trained over 2013 annotated utterances. Once the question has been typed, QA proceeds to focus detection also using machine learning techniques (Mikhailsian et al., 2009). Detected foci include possibly anaphoric expressions ("Who was he?", "Tell me more about the castle"). These expressions are resolved against the dialogue history and geographical context. QA then proceeds to a textual search on texts from the Gazetteer of Scotland (Gittings, 2012) and Wikipedia, and definitions from WordNet glosses. The task is similar to TAC KBP 2013 Entity Linking Track and named en-

tity disambiguation (Cucerzan, 2007). Candidate answers are reranked using a trained confidence score with the top candidate used as the final answer. These are usually long, descriptive answers and are provided as a flow of sentence chunks that the user can interrupt (see table 2). The Interaction Manager queries the QA model and pushes information when a salient PoI is in the vicinity of the user.

> *"Edinburgh's most famous and historic thoroughfare, which has formed the heart of the Old Town since mediaeval times. The Royal Mile includes Castlehill, the Lawnmarket, the Canongate and the Abbey Strand, but, is officially known simply as the High Street."*

Table 2: QA output: query on "Royal Mile"

3.5 Mobile client

The mobile client app, installed on an Android smartphone (Samsung Galaxy S3), connects the user to the dialogue system using a 3G data connection. The client senses the user's location using positioning technology using GNSS satellites (GPS and GLONASS) which is sent to the dialogue system at the rate of one update every two seconds. It also sends pace rate of the user from the accelerometer sensor. In parallel, the client also places a phone call using which the user communicates with the dialogue system.

4 Baseline system

The baseline system chosen for evaluation was Samsung S-Voice, a state-of-the-art commercial smartphone speech interface. S-Voice is a Samsung Android mobile phone app that allows a user to use the functionalities of device using a speech interface. For example, the user can say "Call John" and it will dial John from the user's contacts. It launches the Google Navigation app when users request directions and it activates Google Search for open ended touristic information questions. The Navigation app is capable of providing instructions in-situ using speech. We used the S-Voice system for comparison because it provided an integrated state-of-the-art interface to use both a navigation app and also an information-seeking app using the same speech interface. Users were encouraged to use these apps using speech but were allowed to use the GUI interface when using speech wasn't working (e.g. misrecognition of local names). Users obtained the same kind of in-

formation (i.e. navigation directions, descriptions about entities such as people, places, etc) from the baseline system as they would from our system. However, our system interacted with the user using the speech modality only.

5 Experimental design

Spacebook and the baseline were evaluated in the summer of 2012. We evaluated both systems with 17 subjects in the streets of Edinburgh. There were 11 young subjects (between 20 and 26 years, mean=22 ± 2) and 6 older subjects (between 50 and 71 years, mean=61 ± 11). They were mostly native English speakers (88%). 59% of the users were regular smartphone users and their mean overall time spent in the city was 76 months. The test subjects had no previous experience with the proposed system. They were recruited via email adverts and mail shots. Subjects were given a task sheet with 8 tasks in two legs (4 tasks per leg). These tasks included both navigation and tourist information tasks (see table 3). Subjects used our system for one of the legs and the baseline system for the other and the order was balanced. Each leg took up to 30 mins to finish and the total duration including questionnaires was about 1.5 hours. Figure 2 shows the route taken by the subjects. The route is about 1.3 miles long. Subjects were followed by the evaluator who made notes on their behaviour (e.g. subject looks confused, subject looks at or manipulates the phone, subject looks around, etc).

Subjects filled in a demographic questionnaire prior to the experiment. After each leg, they filled in a system questionnaire (see appendix) rating their experience. After the end of the experiment, they filled out a comparative questionnaire and were debriefed. They were optionally asked to elaborate on their questionnaire ratings. Users were paid £20 after the experiment was over.

6 Results

Subjects were asked to identify tasks that they thought were successfully completed. The perceived task success rates of the two systems were compared for each task using the Chi square test. The results show that there is no statistically significant difference between the two systems in terms of perceived task success although the baseline system had a better task completion rate in tasks 1-3, 5 and 6. Our system performed better in
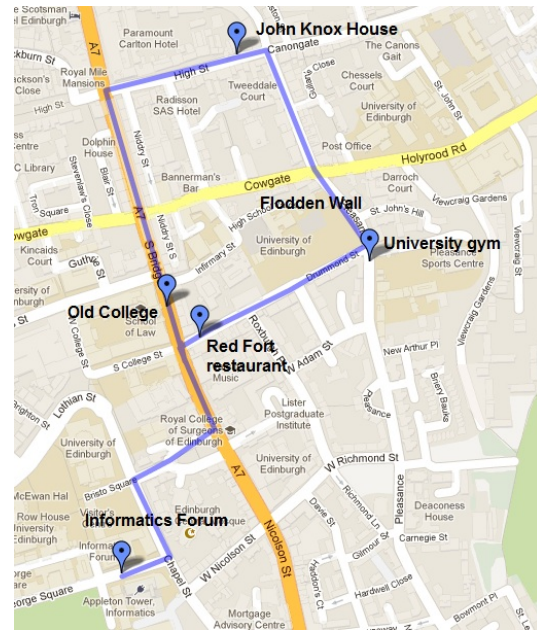

Figure 2: Task route

tourist information tasks (4, 7) (see table 4).

Task	Our system	Baseline	p
T1 (N)	77.7	100	0.5058
T2 (TI)	88.8	100	0.9516
T3 (N)	100	100	NA
T4 (TI)	100	87.5	0.9516
T5 (N+TI)	62.5	100	0.1654
T6 (N+TI)	87.5	100	0.9516
T7 (TI)	100	55.5	0.2926
T8 (N)	75.0	88.8	0.9105

Table 4: % Perceived Task success - task wise comparison (N - navigation task, TI - Tourist Information task)

The system questionnaires that were filled out by users after each leg were analysed. These consisted of questions concerning each system to be rated on a six point Likert scale (1-Strongly Disagree, 2-Disagree, 3-Somewhat Disagree, 4-Somewhat Agree, 5-Agree, 6-Strongly Agree). The responses were paired and tested using a Wilcoxon Sign Rank test. Median and Mode for each system and significance in differences are shown in table 5. Results show that although our system is not performing significantly better than the baseline system (SQ1-SQ10 except SQ7), users seem to find it more understanding (SQ7) and more interesting to interact with (SQ11) than the baseline. We grouped the subjects by age group and tested their responses. We found that the young subjects (age group 20-26), also felt that

Leg 1
(Task 1) Ask the system to guide you to the Red Fort restaurant.
(Task 2) You've heard that Mary Queen of Scots lived in Edinburgh. Find out about her.
(Task 3) Walk to the university gym.
(Task 4) Near the gym there is an ancient wall with a sign saying "Flodden Wall". Find out what that is.
Leg 2
(Task 5) Try to find John Knox House and learn about the man.
(Task 6) Ask the system to guide you to the Old College. What can you learn about this building?
(Task 7) Try to find out more about famous Edinburgh people and places, for example, David Hume, John Napier, and Ian Rankin. Try to find information about people and places that you are personally interested in or that are related to what you see around you.
(Task 8) Ask the system to guide you back to the Informatics Forum.

Table 3: Tasks for the user

they learned something new about the city using it (SQ12) ($p < 0.05$) while the elderly (age group 50-71) didn't. We also found statistically significant differences in smartphone users rating for our system on their learning compared to the baseline (SQ12).

Subjects were also asked to choose between the two systems given a number of requirements such as ease of use, use for navigation, tourist information, etc. There was an option to rank the systems equally (i.e. a tie). They were presented with the same requirements as the system questionnaire with one additional question - "Overall which system do you prefer?" (CQ0). Users' choice of system based on a variety of requirements is shown in table 6. Users' choice counts were tested using Chi-square test. Significant differences were found in users' choice of system for navigation and tourist information requirements. Users preferred the baseline system for navigation (CQ2) and our system for touristic information (CQ3) on the city. Although there was a clear choice of systems based on the two tasks, there was no significant preference of one system over the other overall (CQ0). They chose our system as the most interesting system to interact with (CQ11) and that it was more informative than the baseline (CQ12). Figure 3 shows the relative frequency between user choices on comparative questions.

7 Analysis

Users found it somewhat difficult to navigate using Spacebook (see comments in table 7). Although the perceived task success shows that our system was able to get the users to their destination and there was no significant difference between the two systems based on their questionnaire response on navigation, they pointed out a number of issues and suggested a number of modifications. Many
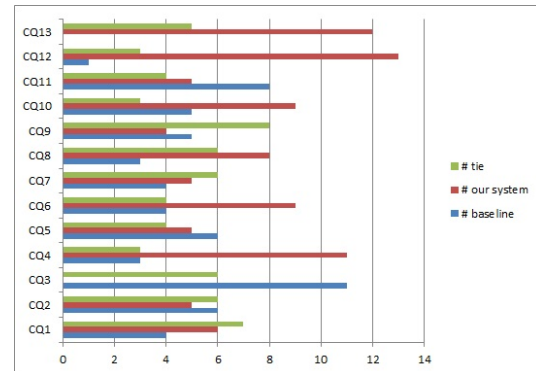

Figure 3: Responses to comparative questions

users noted that a visual map and the directional arrow in the baseline system was helpful for navigation. In addition, they noted that our system's navigation instructions were sometimes not satisfactory. They observed that there weren't enough instructions coming from the system at street junctions. They needed more confirmatory utterances (that they are walking in the right direction) (5 users) and quicker recovery and notification when walking the wrong way (5 users). They observed that the use of street names was confusing sometimes. Some users also wanted a route summary before the navigation instructions are given.

The problem with Spacebook's navigation policy was that it did not, for example, direct the user via easily visible landmarks (e.g. "Head towards the Castle"), and relies too much on street names. Also, due to the latency in receiving GPS information, the IM sometimes did not present instructions soon enough during evaluation. Sometimes it received erroneous GPS information and therefore got the user's orientation wrong. These problems will be addressed in the future version. Some users did find navigation instructions useful because of the use of proximal landmarks such

Question	B Mode	B Median	S Mode	S Median	p
SQ1 - Ease of use	4	4	5	4	0.8207
SQ2 - Navigation	4	4	5	4	0.9039
SQ3 - Tourist Information	2	3	4	4	0.07323
SQ4 - Easy to understand	5	5	5	5	0.7201
SQ5 - Useful messages	5	4	5	4	1
SQ6 - Response time	5	5	2	2	0.2283
SQ7 - Understanding	3	3	**5**	**4**	**0.02546**
SQ8 - Repetitive	2	3	2	3	0.3205
SQ9 - Aware of user environment	5	5	4	4	0.9745
SQ10 - Cues for guidance	5	5	5	5	0.1371
SQ11 - Interesting to interact with	5	4	**5**	**5**	**0.01799**
SQ12 - Learned something new	5	4	5	5	0.08942

Table 5: System questionnaire responses (B=Baseline, S=our system)

Task	Baseline Preferred	Our system Preferred	Tie	p-value
CQ0	23.52	35.29	41.17	0.66
CQ1	35.29	29.41	35.29	0.9429
CQ2	**64.70**	0	35.29	**0.004**
CQ3	17.64	**64.70**	17.64	**0.0232**
CQ4	35.29	29.41	23.52	0.8187
CQ5	23.52	52.94	23.52	0.2298
CQ6	23.52	29.41	35.29	0.8187
CQ7	17.64	47.05	35.29	0.327
CQ8	29.41	23.52	47.05	0.4655
CQ9	29.41	52.94	17.64	0.1926
CQ10	47.05	29.41	23.52	0.4655
CQ11	5.88	**76.47**	17.64	**0.0006**
CQ12	0	**70.58**	29.41	**0.005**

Table 6: User's choice on comparative questions (CQ are the same questions as SQ but requesting a ranking of the 2 systems)

as KFC, Tesco, etc. (popular chain stores). Some users also suggested that our system should have a map and that routes taken should be plotted on them for reference. Based on the ratings and observations made by the users, we conclude that our first hypothesis that Spacebook would be more efficient for navigation than the baseline because of its speech-only interface was inconclusive. We believe so because users' poor ratings for Spacebook may be due to the current choice of dialogue policy for navigation. It may be possible to reassure the user with a better dialogue policy with just the speech interface. However, this needs further investigation.

Users found the information-search task interesting and informative when they used Spacebook (see sample user comments in table 8). They also found push information on nearby PoIs unexpected and interesting as they would not have found them otherwise. Many users believed that this could be an interesting feature that could help tourists. They also found that asking questions and

finding answers was much easier with Spacebook compared to the baseline system, where sometimes users needed to type search keywords in. Another user observation was that they did not have to stop to listen to information presented by our system (as it was in speech) and could carry on walking. However, with the baseline system, they had to stop to read information off the screen. Although users in general liked the QA feature, many complained that Spacebook spoke too quickly when it was presenting answers. Some users felt that the system might lose context of the navigation task if presented with a PoI question. In contrast, some others noted Spacebook's ability to interleave the two tasks and found it to be an advantage.

Users' enthusiasm for our system was observed when (apart from the points of interest that were in the experimental task list) they also asked spontaneous questions about James Watt, the Talbot Rice gallery, the Scottish Parliament and Edinburgh Castle. Some of the PoIs that the system pushed information about were the Royal College of Surgeons, the Flodden Wall, the Museum of Childhood, and the Scottish Storytelling Centre. Our system answered a mean of 2.5 out of 6.55 questions asked by users in leg 1 and 4.88 out of 8.5 questions in leg 2. Please note that an utterance is sent to QA if it is not parsed by the parser and therefore some utterances may not be legitimate questions themselves. Users were pushed a mean of 2.88 and 6.37 PoIs during legs 1 and 2. There were a total of 17 "tell me more" requests requesting the system to present more information (mean=1.35 ± 1.57).

Evaluators who followed the subjects noted that the subjects felt difficulty using the baseline system as they sometimes struggled to see the screen

1. "It's useful when it says 'Keep walking' but it should say it more often."
2. "[Your system] not having a map, it was sometimes difficult to check how aware it was of my environment."
3. "[Google] seemed to be easier to follow as you have a map as well to help."
4. "It told me I had the bank and Kentucky Fried Chicken so I crossed the road because I knew it'd be somewhere over beside them. I thought 'OK, great. I'm going the right way.' but then it didn't say anything else. I like those kind of directions because when it said to go down Nicolson Street I was looking around trying to find a street sign."
5. "The system keeps saying 'when we come to a junction, I will tell you where to go', but I passed junctions and it didn't say anything. It should say 'when you need to change direction, I will tell you.'"
6. "I had to stop most of the times for the system to be aware of my position. If walking very slowly, its awareness of both landmarks and streets is excellent."

Table 7: Sample user comments on the navigation task

1. "Google doesn't *offer* any information. I would have to know what to ask for..."
2. "Since many information is given without being asked for (by your system), one can discover new places and landmarks even if he lives in the city. Great feature!!"
3. "I didn't feel confident to ask [your system] a question and still feel it would remember my directions"
4. "Google could only do one thing at a time, you couldn't find directions for a place whilst learning more."
5. "If she talked a little bit slower [I would use the system for touristic purposes]. She just throws masses of information really, really quickly."

Table 8: Sample user comments on the tourist information task

in bright sunlight. They sometimes had difficulty identifying which way to go based on the route plotted on the map. In comparison, subjects did not have to look at the screen when they used our system. Based on the ratings and observations made by the users about our system's tourist information features such as answering questions and pushing PoI information, we have support for our second hypothesis: that users find a dialogue interface which integrates question-answering and navigation within a shared context to be useful for finding information about entities in the urban environment.

8 Future plans

We plan to extend Spacebook's capabilities to address other challenges in pedestrian navigation and tourist information. Many studies have shown that visible landmarks provide better cues for navigation than street names (Ashweeni and Steed, 2006; Hiley et al., 2008). We will use visible landmarks identified using the visibility engine to make navigation instructions more effective, and we plan to include entities in dialogue and visual context as candidates for PoI push, and to implement an adaptive strategy that will estimate user interests and push information that is of interest to them. We are also taking advantage of user's local knowledge of the city to present navigation instructions only for the part of the route that the user does not have any knowledge of. These features, we believe, will make users' experience of the interface more pleasant, useful and informative.

9 Conclusion

We presented a mobile dialogue app called Spacebook to support pedestrian users in navigation and tourist information gathering in urban environments. The system is a speech-only interface and addresses navigation and tourist information in an integrated way, using a shared dialogue context. For example, using the navigational context, Spacebook can push point-of-interest information which can then initiate touristic exploration tasks using the QA module.

We evaluated the system against a state-of-the-art baseline (Samsung S-Voice with Google Navigation and Search) with a group of 17 users in the streets of Edinburgh. We found that users found Spacebook interesting to interact with, and that it was their system of choice for touristic information exploration tasks. These results were statistically significant. Based on observations and user ratings, we conclude that our speech-only system was less preferred for navigation and more preferred for tourist information tasks due to features such as PoI pushing and the integrated QA module, when compared to the baseline system. Younger users, who used Spacebook, even felt that they learned new facts about the city.

Acknowledgments

References

K. B. Ashweeni and A. Steed. 2006. A natural wayfinding exploiting photos in pedestrian navigation systems. In *Proceedings of the 8th conference on Human-computer interaction with mobile devices and services*.

P. Bartie and W. Mackaness. 2006. Development of a speech-based augmented reality system to support exploration of cityscape. *Transactions in GIS*, 10:63–86.

P. Bartie and W. Mackaness. 2012. D3.4 Pedestrian Position Tracker. Technical report, The SPACE-BOOK Project (FP7/2011-2014 grant agreement no. 270019).

P. Bartie and W. Mackaness. 2013. D3.1.2 The Space-Book City Model. Technical report, The SPACE-BOOK Project (FP7/2011-2014 grant agreement no. 270019).

D. Byron, A. Koller, J. Oberlander, L. Stoia, and K. Striegnitz. 2007. Generating Instructions in Virtual Environments (GIVE): A challenge and evaluation testbed for NLG. In *Proceedings of the Workshop on Shared Tasks and Comparative Evaluation in Natural Language Generation*.

S. Cucerzan. 2007. Large-scale named entity disambiguation based on Wikipedia data. In *Proceedings of EMNLP-CoNLL*.

R. Dale, S. Geldof, and J. Prost. 2003. CORAL : Using Natural Language Generation for Navigational Assistance. In *Proceedings of ACSC2003, Australia*.

Nina Dethlefs and Heriberto Cuayáhuitl. 2011. Hierarchical Reinforcement Learning and Hidden Markov Models for Task-Oriented Natural Language Generation. In *Proc. of ACL*.

B. Gittings. 2012. The Gazetteer for Scotland - http://www.scottish-places.info.

H. Hiley, R. Vedantham, G. Cuellar, A. Liuy, N. Gelfand, R. Grzeszczuk, and G. Borriello. 2008. Landmark-based pedestrian navigation from collections of geotagged photos. In *Proceedings of the 7th Int. Conf. on Mobile and Ubiquitous Multimedia (MUM)*.

S. Janarthanam and O. Lemon. 2011. The GRUVE Challenge: Generating Routes under Uncertainty in Virtual Environments. In *Proceedings of ENLG*.

S. Janarthanam, O. Lemon, X. Liu, P. Bartie, W. Mackaness, T. Dalmas, and J. Goetze. 2012. Integrating location, visibility, and Question-Answering in a spoken dialogue system for Pedestrian City Exploration. In *Proc. of SIGDIAL 2012, S. Korea*.

H. Kashioka, T. Misu, E. Mizukami, Y. Shiga, K. Kayama, C. Hori, and H. Kawai. 2011. Multimodal Dialog System for Kyoto Sightseeing Guide. In *Asia-Pacific Signal and Information Processing Association Annual Summit and Conference*.

S.S. Keerthi, S. K. Shevade, C. Bhattacharyya, and K. R. K. Murthy. 1999. Improvements to Platt's SMO Algorithm for SVM Classifier Design. *Neural Computation*, 3:637–649.

J. Ko, F. Murase, T. Mitamura, E. Nyberg, M. Tateishi, I. Akahori, and N. Hataoka. 2005. CAMMIA: A Context-Aware Spoken Dialog System for Mobile Environments. In *IEEE ASRU Workshop*.

C. Kray, K. Laakso, C. Elting, and V. Coors. 2003. Presenting Route Instructions on Mobile Devices. In *Proceedings of IUI 03, Florida*.

R. Malaka and A. Zipf. 2000. Deep Map - challenging IT research in the framework of a tourist information system. In *Information and Communication Technologies in Tourism 2000*, pages 15–27. Springer.

A. Mikhailsian, T. Dalmas, and R. Pinchuk. 2009. Learning foci for question answering over topic maps. In *Proceedings of ACL 2009*.

D. Montello. 1993. Scale and multiple psychologies of space. In A. U. Frank and I. Campari, editors, *Spatial information theory: A theoretical basis for GIS*.

M. Raubal and S. Winter. 2002. Enriching wayfinding instructions with local landmarks. In *Second International Conference GIScience. Springer, USA*.

C.J. Shroder, W. Mackaness, and B. Gittings. 2011. Giving the Right Route Directions: The Requirements for Pedestrian Navigation Systems. *Transactions in GIS*, pages 419–438.

N. Webb and B. Webber. 2009. Special Issue on Interactive Question Answering: Introduction. *Natural Language Engineering*, 15(1):1–8.

P. A. Zandbergen and S. J. Barbeau. 2011. Positional Accuracy of Assisted GPS Data from High-Sensitivity GPS-enabled Mobile Phones. *Journal of Navigation*, 64(3):381–399.